THE IMPACT OF SOCIO-ECONOMIC BACKGROUND AND BIRTH ATTRIBUTES ON INFANT MORTALITY IN BANGLADESH: A CROSS-SECTIONAL STUDY

A PROJECT REPORT SUBMITTED IN PARTIAL FULFILLMENT FOR THE REQUIREMNET OF MASTER OF SCIENCE IN STATISTICS AT HAJEE MOHAMMAD DANESH SCIENCE AND TECHNOLOGY UNIVERSITY, DINAJPUR-5200

DEPARTMENT OF STATISTICS

HAJEE MOHAMMAD DANESH SCIENCE AND TECHNOLOGY UNIVERSITY, DINAJPUR-5200

JULY-DECEMBER, 2019

THE IMPACT OF SOCIO-ECONOMIC BACKGROUND AND BIRTH ATTRIBUTES ON INFANT MORTALITY IN BANGLADESH: A CROSS-SECTIONAL STUDY

BY

MD. SHOHIDULLAH SAMRAT

STUDENT ID: 19051265

AUTHORIZED AS TO THE MODE AND SUBJECT MATTER BY

(DR. RAJIB DEY)	(MD. SOBUJ ALI)
ASSISTANT PROFESSOR	ASSISTANT PROFESSOR
SUPERVISOR	CO-SUPERVISOR

(PROF. DR. MD. EARFAN ALI KHONDAKER)

CHAIRMAN

DEPARTMENT OF STATISTICS

HAJEE MOHAMMAD DANESH SCIENCE AND TECHNOLOGY UNIVERSITY, DINAJPUR-5200

JULY-DECEMBER, 2019

Acknowledgements

All praise is due to Allah, the beneficent and the merciful. I express my sacred gratitude to the omnipotent and omniscient Allah. This project would not have been possible without the wonderful assistance and direction of some individuals. I would like to convey my appreciation to Dr. Rajib Dey, Assistant professor and Md. Sobuj Ali, Assistant professor, Department of Statistics, Hajee Mohammad Danesh Science & Technology University (HSTU), for mentoring me throughout this study. Without their encouragement and motivation, I would not have been able to produce this project. I'd want to convey my sincere appreciation to the Department of Statistics at HSTU in particular for giving me an amazing opportunity to keep learning. I'd also like to thank my friends for their enthusiasm and unique insights on the project. Last but not least, I am grateful to my parents for their everlasting belief in my accomplishments, as well as their endless love, encouragement, and support.

Abstract

Sustainable Development Goals are to reduce newborn mortality (NM) and under-five mortality (U5M) rates to 12 and 25 deaths per 1,000 live births by 2030. Following the successful implementation of the Millennium Development Goal (MDG), Bangladesh, a lower-middle-income nation in South Asia, had a considerable drop in U5M (from 133 to 30.2 fatalities per 1000 live births between 1990 and 2018). However, rates remain high in South Asian nations, with Bangladesh ranking third after Pakistan (74 per 1000 live births) and India (38.8 per 1000 live births). There is no debate that further work and efforts are required to ensure further NM and U5M reductions to meet the corresponding SDG objectives. It is evident from other Bangladesh Demographic and Health Survey (BDHS) studies that infant mortality in Bangladesh is decreasing day by day. That is due to several socio-economic and health-related factors that influence the maternal and newborn baby's health. We have worked in this study based on BDHS data (2017-18) having the number of total data of 20,127 households having 47,828 children with Division, Type of place of residence, Wealth index combined, Mother's educational level, Age of mother at birth, Sex of child, Birth order of the child, preceding birth interval, successive birth interval, etc. variables. To examine the factors that influence infant mortality, a Binary Logistic Regression and a Random Forest Regression analysis were performed. The findings of the Binary Logistic Regression were more suitable, and the odds ratio was compared with confidence intervals to assist in the interpretation of the data. According to this study, babies in Bangladesh are at greater risk of death due to societal factors such as mothers' lack of formal education and poverty. Infant mortality is more likely to occur if the mother is older and born higher in the pecking order. Preventing infant mortality requires community-based programs, emphasizing improving family-based healthcare for newborns.

Keyword:

BDHS, Infant Mortality, Chi-Square Test, Binary Logistic Regression, Random Forest Regression, SDG

Table of contents

List of Tables

List of Figures

Chapter 1 Introduction

Bangladesh's neonatal and infant mortality rates have been declining. According to the Bangladesh Demographic and Health Survey (BDHS) study, the under-5 mortality rate in Bangladesh was 45 fatalities per 1,000 live births in the five years preceding the study (about 2014-2018). (BDHS 2017-18). The newborn mortality rate was 38 per 1,000 live births, while the child mortality rate was 7 per 1,000. The risk of dying in the first month of life (30 deaths per 1,000 live births) is approximately four times that of dying in the next 11 months (8 deaths per 1,000 live births). It's also worth mentioning that 67 percent of all deaths in children under five occur during the newborn period. Under-5 mortality was 40 deaths per 1,000 live births, newborn mortality was 34 deaths per 1,000 live births, and neonatal mortality was 26 deaths per 1,000 live births, according to the 2019 MICS survey of the Bangladesh Bureau of Statistics (BBS). Between 1993-1994 and 2017-18, under-5, infant and neonatal mortality decreased by 66%, 57%, and 42%, respectively. However, reductions were unusually slow or negligible from 2014 to 2017-18. Previous research has found a variety of risk factors for early childhood death, including biodemographic and sociodemographic variables based on the mother and child's features and the circumstances of the birth. The analysis includes biodemographic indicators such as the sex of the child, the mother's age at birth, birth order, and previous birth interval. Place of residence, division, mother's schooling, and wealth index were among the sociodemographic parameters taken under consideration.

The purpose of this study is to explore the biodemographic and sociodemographic factors that influence infant mortality. In order to explore the causes and consequences of significant dispersion in infant mortality, the study was conducted. We will utilize secondary data from the BDHS for the 2017-18 period.

To determine the attributes that contribute to infant mortality dispersion, a Binary Logistic Regression and a Random Forest Regression analysis were done on the number of newborn fatalities to determine the factors that contribute to infant mortality spreading.

1.1 Background of the study:

Socioeconomic inequities in health and mortality have long been a matter of attention for our society, regardless of profound influence, and Bangladesh is no exception. Though health services, including child health care, are provided by the public sector for free in many countries, poor people utilize them in smaller percentages than better-off people because they are less educated and face cultural and societal challenges in accessing care. Discrimination in health concerns has recently received increasing worldwide attention, with specific reference as development goals in global agendas such as the Millennium Development Goals (MDGs) and Sustainable Development Goals (SDGs). The Commission on Social Determinants of Health has been helpful to promote dialogue and promoting worldwide policies and strategic goals. What was left unknown was how best to minimize the inequalities to an acceptable level. While developing countries, such as Bangladesh, have achieved considerable progress in improving health and decreasing death rates, health but mortality inequalities remain the same. Though Bangladesh has made headway in lowering urban-rural and regional differences in under-five mortality, but socioeconomic disparities exist. Demographic analyses were conducted using SPSS v25, as well as the Logistic model and Random forest models were carried out in R version 4.1.3.

Many poverty alleviation programs, particularly microfinance, targeting the most underprivileged in Bangladesh have been undertaken to improve the socioeconomic status of the underprivileged and eradicate socio-economic imbalances. The impact of such development programs on decreasing socioeconomic inequalities has been investigated and shown to reduce health disparity. Along with microfinance, Bangladesh launched a free universal primary education program with a focus on females, funded by cash/material incentives for education, leading to a major increase in female schooling. The growth of the public sector's free vaccination program, family planning services, and basic healthcare to bring services to people's doorsteps, particularly in rural regions, resulted in near-universal vaccine coverage and increasing use of contraceptives.

Mothers' education, one of several non-health determinants determining child survival, has been proven to be highly significant across nations, particularly in contexts with poor health facilities and socioeconomic development, also including Bangladesh.

Among the socio-economic parameters stated above, sex of children, birth order of children, mothers' age at birth, household socioeconomic status such as wealth index, and site of residence have been determined to be essential.

Given the relevance of mother's education and the socioeconomic disparity in child mortality, we evaluated the trajectory of socioeconomic inequality in infant mortality in rural Bangladesh using data from the BDHS 2017-18 dataset. We also looked at whether the mother's education and family financial position had an impact on minimizing social inequities.

1.2 Motivation of the study

This study investigates trends and estimates the Infant mortality rate (IMR) in Bangladesh, as well as inspects Socio-Economic behaviors and certain birth characteristics.

1.3 Objectives of the study

- To investigate the biodemographic variables influencing infant mortality in Bangladesh.
- To determine the sociodemographic determinants that affect infant mortality in Bangladesh.
- To compare the use of Binary Logistic Regression and Random Forest Regression on BDHS 2017-18 data with significant dispersion.

1.4 Outline of the study

Chapter 2 reviews ideas about methods and materials. A logistic regression model and random forest regressor model were used to study the characteristics that influence infant mortality in Bangladesh. The data were examined with the R program, SPSS 25 statistical program, and Microsoft Excel 2016.

Chapter 3 reviews ideas about the BDHS dataset and the study variables used for the study. The chapter also includes the ethical approval process and casual diagram of the study.

Chapter 4 is related to Analysis, results, and discussion. In this section of the research, we check out the factors that influence infant mortality. However, Data from the BDHS from 2017-2018 were reviewed for multivariate analysis.

Chapter 5 is related to the summary and conclusion. In this section of the study, it is recommended that not only the government but also the common people take necessary steps to reduce infant mortality in Bangladesh. In this chapter, we also came to our findings. Using data from the Bangladesh Demographic and Health Surveys, this study explores the situation and variables.

In the Reference section, the reference sources are mentioned.

Chapter 2 Methodology

2.1 Introduction

Data from the Bangladesh Demographic and Health Survey was used in this study (2017-18). To describe an ever had infant mortality, the Chi-square test and logistic regression analysis were used. Chi-square test was used in this study to examine the determinants of infant mortality. Logistic regression analysis was applied to find out the most significant factors of infant mortality. It was found that the covariates were education level, Respondent Currently Working, Type of place of Residence, Wealth index combined, Religion, Highest Educational Level, Division the significant effect on infant mortality.

2.2 Analytical approach

Let, Y be the outcome variable which has two levels and X is another random variable which is our explanatory variables and also a binary random variables.

2.3 Chi-square test statistic

The formula for calculating the chi-square test statistic is:

$$\chi^2 = \frac{\sum(\boldsymbol{O} - \boldsymbol{E})^2}{\boldsymbol{E}}$$

Where,

O represents the observed frequency,

E represents the expected frequency under the null hypothesis,

Which is computed by:

E = (row total × column total) / sample size

We will compare the value of the test statistic to the critical value of $\chi^{2\alpha}$ with degree of freedom (r - 1) (c - 1) and reject the null hypothesis if $\chi^2 > \chi^{2\alpha}$.

2.4 Test hypothesis

The higher the chi-square value, the more likely it is to be significant, allowing the null hypothesis to be rejected and the variables to be linked. Because the p-value in the test result is smaller than the average alpha value, the null hypothesis that all variables are independent of one another is rejected.

2.5 P-value

The smallest level of significance is the p-value. The likelihood of receiving a test statistic result as severe as or more extreme than the one actually computed is the p-value for a hypothesis test. A statistic's p-value is the minimum number at which the null hypothesis may be rejected. We reject the null hypothesis if the p-value is higher than the significance level. We don't reject if the p-value is smaller than the significance level.

2.6 Logistic regression model

2.6.1 Interpreting parameters in logistic regression

For a binary response variable Y and an explanatory variable X,

$$\pi(x) = P(X = x) = 1 - P(X = x).$$

The logistic regression model is

$$\pi(x) = \frac{expexp\,(\alpha+\beta x)}{1+expexp\,(\alpha+\beta x)} \qquad (1)$$

Equivalently, the log odds, called the *logit*, has the linear relationship

$$logit[\pi(x)] = log\,\frac{\pi(x)}{1-\pi(x)} = \alpha + \beta x \qquad (2)$$

This equates the logit link function to the linear predictor.

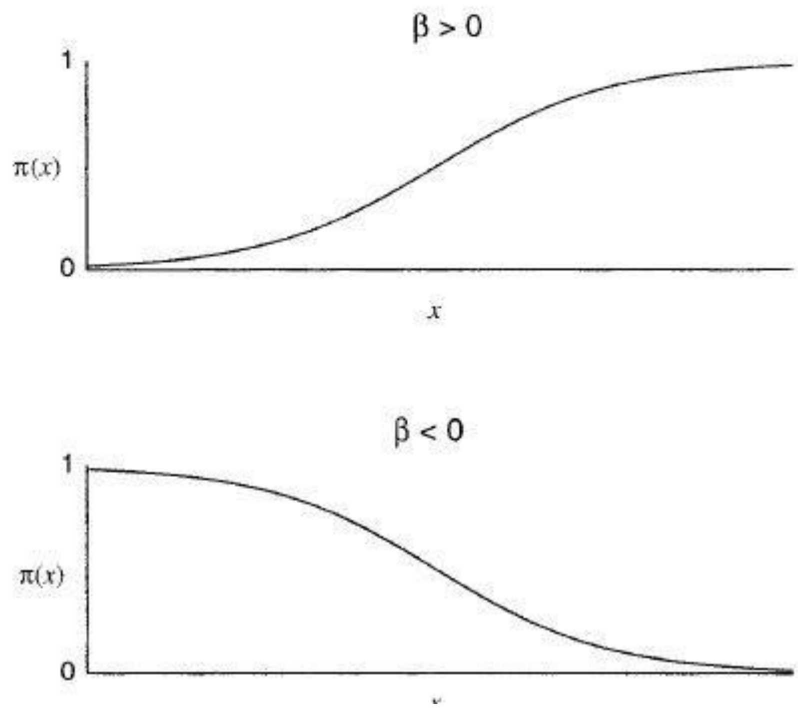

Figure 2.1 Logistic regression function

2.7 Interpreting β

Its sign determines whether $\pi(x)$ is increasing or decreasing as x increases. The rate of climb or descent increases as $|\beta|$ increases as $\beta \to 0$ the curve flattens to a horizontal straight line. When $\beta = 0$, Y is independent of X. For quantitative x with $\beta > 0$, the curve for $\pi(x)$ has the shape of the cdf of the logistic distribution. Since the logistic density is symmetric, $\pi(x)$ approaches 1 at the same rate that it approaches 0.

Taking exponentiation both sides of (2) shows that the odds are an exponential function of x. This provides a basic interpretation for the magnitude of β : The odds increase multiplicatively by e^{β} for every 1-unit increase in x. In other words, e^{β} is an odds ratio, the odds at X = x+1 divided by the odds at X = x.

2.7.1 Binomial GLM for 2 × 2 contingency tables

Among the simplest GLMs for a binary response is the one having a single explanatory variable X that is also binary. Label its values by 0 and 1. For a given link function, the GLM

$$link[\pi(x)] = \alpha + \beta x$$

has the effect of X described by

$$\beta = link[\pi(1)] - lin\,k[\pi(0)]$$

For the identity link, $\beta = \pi(1) - \pi(0)$ is the difference between proportions. For the log link, $\beta = log[\pi(1)] - log[\pi(0)] = log[\pi(1)/\pi(0)]$ is the log relative risk. For the logit link,

$$\beta = logit[\pi(1)] - logit[\pi(0)] = log\frac{\pi(1)}{1-\pi(1)} - log\frac{\pi(0)}{1-\pi(0)} = log\frac{\pi(1)/(1-\pi(1))}{\pi(0)/(1-\pi(0))}$$

is the log odds ratio. Measures of association for 2 × 2 tables are effect parameters in GLMs for binary data.

Likelihood equation for logistic model

$$\sum_{0}^{i} n_i((y_i - \pi(x_i))x_i = 0$$

This is the likelihood equation for logistic model by solving this equation we can estimate α and β.

2.7.2 Inference for binomial logistic

Deviance for binomial models

The deviance equals

$$2\sum_{0}^{i} n_i y_i log\frac{n_i y_i}{n_i \hat{\pi}_i} + 2\sum_{0}^{i}(n_i - n_i y_i)log\frac{n_i - n_i y_i}{n_i - n_j \hat{\pi}_i}$$

At setting i,

$n_i y_i$ is the number of successes and

$n_i - n_i y_i$ is the number of failures, is $1, \dots, N$.

2.8 Estimation of odds ratio

The fundamental goal of estimating independent variable parameters in logistic regression analysis is to interpret the relationship between the result and the independent variables. The odds ratio is used to make practical inferences from the estimated coefficients of a fitted model. Because the dependent variable is on a continuous scale, the estimated coefficients for the independent variables indicate the slope or rate of change in the function of the dependent variable for every unit change in the independent variable. However, because the dependent variable in logistic regression is categorical, such an interpretation of the computed coefficients is not valid. Because the logit in the independent variables is continuous and linear, the interpretation will be possible if the change in independent variable is compared to the change in logit. As a result, we can see that interpretation entails determining the functional connection between the dependent variable and the independent variables, as well as defining the independent variables' unit of change.

2.9 Random forest regressor

A random forest regressor is a Meta estimator that fits a number of classifying decision trees on various sub- samples of the dataset and uses averaging to improve the predictive accuracy and control over fitting. A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap and Aggregation, commonly known as bagging. The basic behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees.

Random forest has multiple decision trees as base learning models. We randomly perform row sampling and feature sampling from the dataset forming sample datasets for every model. This part is called bootstrap.

2.10 Akaike information criterion

When several models are available, one can compare the models performance based on several likelihood measures which have been proposed in the statistical literature. One of the most popularly used measures is AIC. The AIC penalized a model with a larger number of parameters and is defined as

$$\boldsymbol{AIC = -2lnL + 2p}$$

Where denotes the fitted log likelihood and p the number of parameters. A relatively small value of AIC is favorable for the fitted model.

An alternative formula for least squares regression type analyses for normally distributed errors:

Where:

$$\boldsymbol{AIC = n\, log\,(2) + 2k}$$

RSS = Residual Sum of Squares/n,

n = sample size,

K is the number of model parameters.

2.11 Source of data

Since this study has its focus on the abortion in Bangladesh. Secondary data BDHS (17-18) was used in this investigation. This study is based on the data of 20127 Bangladeshi women who had ever married. Ever-married women mean they have been married at least once throughout their lives even though their current marital status may not be “married”.

2.12 Statistical analysis

All statistical analyses in this study were performed using the R program (v-4.1.3), SPSS (v-25) and Microsoft Excel 2016. Then logistic regression model is fitted and compared with random forest model.

Chapter 3 Description of variables

3.1 Bangladesh demographic and health survey (BDHS) background

This research is based on data from a country wide cross sectional survey, the Bangladesh demographic and health (DHS) survey 2017-2018. The survey was created to collect detail information on reproductive history, marriage, fertility preferences, family planning methods, breast feeding practices women and children's nutritional status, maternal and child Health line and child mortality, HIV/AIDS knowledge and attitudes and so on.

The sample was weighted in this study using the matching weighting factor supplied for each of the women in the survey data. The first portion will aid in better understanding the current state of child mortality behavior, particularly the childbearing status of married women, and the second will aid in better understanding the factors of infant mortality.

Scientific sampling probability was used to create the survey sample. A thorough birth history was kept, which included all live births conducted by women, as well as their age at first birth, birth date, year of birth, child's sex, and survival status.

An interview was only done if the respondent gave their verbal assent after the interviewer read them and informed consent statement. The DHS data set authorization was obtained from https:// https://dhsprogram.com/data and user guidelines were rigorously afforded to (all DHS data should be made to identify any household or individual respondent interviewed in the survey). Using a data extraction tool, variables were retrieved from the DHS 2017-2018 individual women's dataset.

3.2 Ethical approval

After eliminating all identifiable information from the respondents, this research was best on an examination of existing public domain survey data sets data publicly available online this survey was authorized by Bangladeshi ethics committees and ICF Macro in Calverton New York. The authors obtain permission to download their dataset from this study for the demographic and health survey (DHS) programmed data archivist.

3.3 Casual diagram of the study:

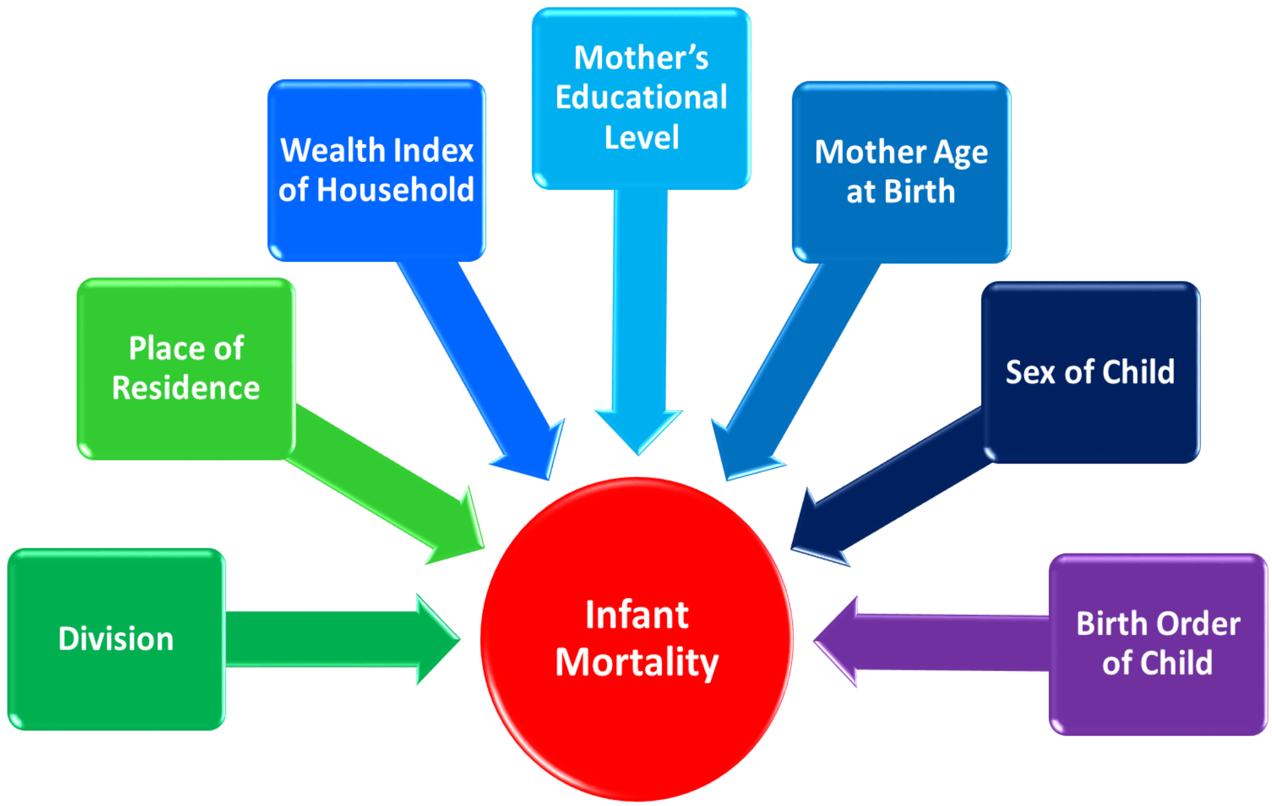

Figure 3.1 The causal diagram of the variables of the study

To determine the influence of demographic, socioeconomic and community service related variables on the number of ANC visits during pregnancy of a woman in Bangladesh, the following covariates are considered. These are

- Division
- Type of place of residence
- Wealth index combined
- Mother's educational level
- Age of mother at birth
- Sex of child
- Birth order of child

3.4 Variables

3.4.1 Response Variable

The key response variable in this study was the age of death of the child form respondents. It has a direct impact on child mortality. The infant mortality influences SDG's third goal of Good Health and Well-being. The third goal of SDG targets by 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births.

The variables of interest for this study were obtained from the individual women questionnaire. The response variable "Age at death (months)" in women individual recode dataset was calculated from 20 repeated measure of the variable "Age at death (months)" from the respondent of the dead child. All dates in the data are expressed in terms of months. Age at death (months) is a continuous variable that ranges from 0 to 48 months of age (Bangladesh Demographic and Health Survey, 2017-18).

The response variable 'age at death' is assessed in months, and the start of exposure to infant mortality is calculated from the variable 'Child is alive'. As a result, the answer variable indicates the status of the child whether the child is dead or alive. The 'age at death (months)' variable indicates their ages at the time of death who were died at different point of time in months.

The research sample consisted of 20127 women aged 10 to 49 who had previously given birth (Bangladesh Demographic and Health Survey, 2017-18). The response variable for this study was the respondent's age at death (months). The child's age at death was recorded into two categories and defined as:

$$Y = \{1, Respondents\ age\ at\ death\ \leq 12;\ 0, otherwise\}$$

The outcome variable of interest in this study is infant mortality. In this case, infant mortality includes child who had either before reaching their first birth or survive first year of birth.

3.4.2 Explanatory variables

As determinants of early motherhood in Bangladesh, several demographics, social, cultural, and geographical characteristics were examined. Education (respondents) was categorized into four groups among socioeconomic variables: no education, primary, secondary, and higher education. The lowest, poorer, medium, richer, and richest categories were used to calculate the wealth index.). As geographical factors, respondents' place of residence (urban vs rural) and division (Barisal, Chattogram, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, and Sylhet) were included. Some birth characteristics like sex of the children, birth order of children and mothers age at birth is also included in the study.

The variables were chosen based on past studies and research expertise to explain the data in the simplest way possible and to minimize redundant predictions. Bangladesh's demographic, social, cultural sectors and birth characteristics are represented by these factors.

3.5 Description of Explanatory variables

The major goal of this research is to figure out what factors influence infant mortality. As a result, we must contend with a plethora of issues related with child mortality. For the sake of this study, certain variables are recorded, and new variables are formed by merging information from many variables rather than using the variables' original codes.

3.5.1 Place of residence

Urbanization and its modernizing effects are expected to reduce infant mortality. Rural women were thought to have a higher infant mortality rate than urban women for the following reasons: urban women have greater access to healthcare than rural women. Infant mortality is higher in rural areas due to hunger and poor health facilities.

3.5.2 Region

The mortality rate of infants varies depending on where they live. A variable called division (Barisal, Chittagong, Dhaka, Khulna, Mymensingh, Rajshahi, Rangpur, and Sylhet) was included in the study to determine the impacts of infant mortality. Different parameters were measured at different locations, exhibiting distinct rates.

3.5.3 Mother's education level

The education of women, in particular, has a significant interaction effect on infant mortality. Education provides new values, aims, and a fresh perspective on life, as well as the capacity to embrace potential ideas. Educated women are more likely than uneducated mothers to provide better care for their children.

3.5.4 Wealth index

The wealth index has an important role in predicting infant mortality rates. This is because the impoverished are unable to offer adequate healthcare for their children and antenatal care for their mothers.

3.5.5 Age of the mother at birth

Mother's age at birth (maternal age) is considered as an influential factor of infant mortality. In this study mother's age at birth was calculated from the mother's age at marriage and the preceding and successive birth intervals. For the simplicity of the analysis, the data were recoded as

20 or less=1
20-29=2
30 or more= 3

3.5.6 Sex of child

Gender of the kid: The gender of the child has a significant impact on infant mortality. Male infants have been shown to have a greater infant mortality rate than female infants on several occasions. Here the multiple responses from some individuals were obtained than for the analysis the data are recoded as

1= female
2= male

3.5.7 Birth order of child

Several studies have discovered that birth order is likely to be linked to infant mortality. Multiple responses from each respondent were collected in this study, and the data were recoded for analytic purposes.

1 = 1
2-3 = 2
4 or more = 3

Chapter 4 Result and Discussion:

4.1 Exploratory statistics

Here the graphical representation of different variables associated with the number Infant mortality:

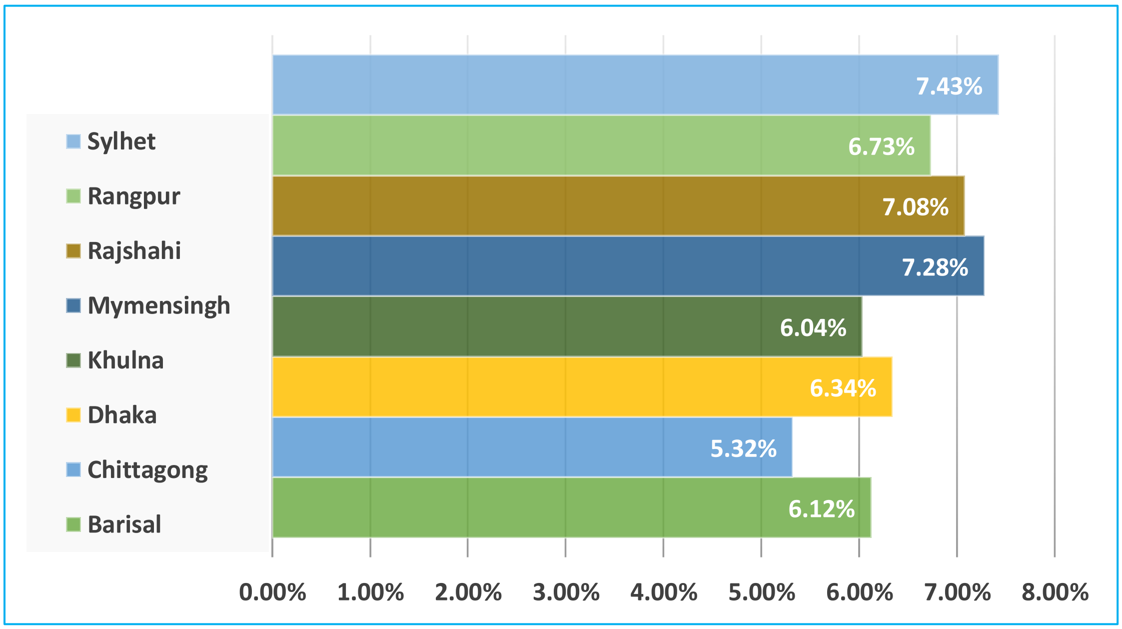

Figure 4.1 Percentage of Infant mortality in different divisions

Figure 4.1 represent that the most percentages of infant mortality occurs in the Sylhet division along with Mymensingh district and the least percentage of infant mortality occur in the Chittagong District.

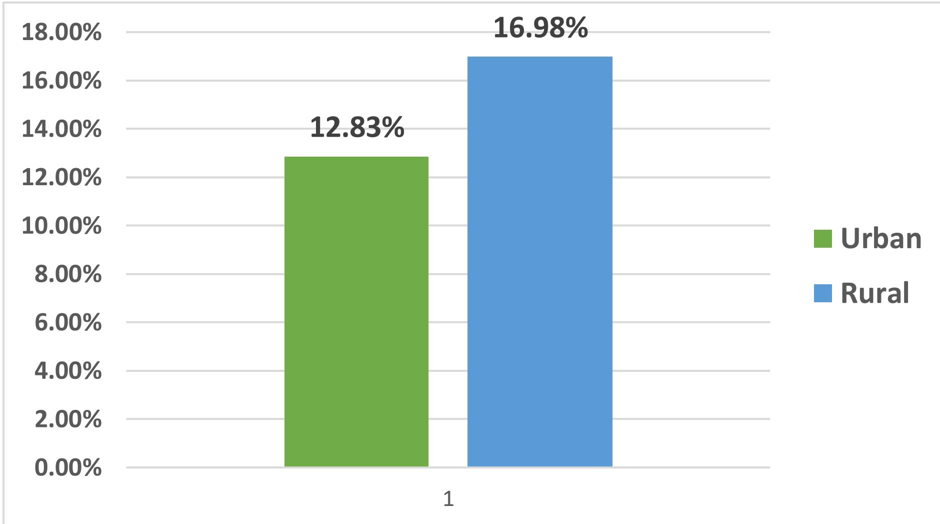

Figure 4.2 Percentage of Infant mortality in different type of place of residence

Figure 4.2 represent that the most percentages of infant mortality occurs in the rural areas and the least percentage of infant mortality occur in the urban areas. Hence the difference of infant mortality in different residence is very high.

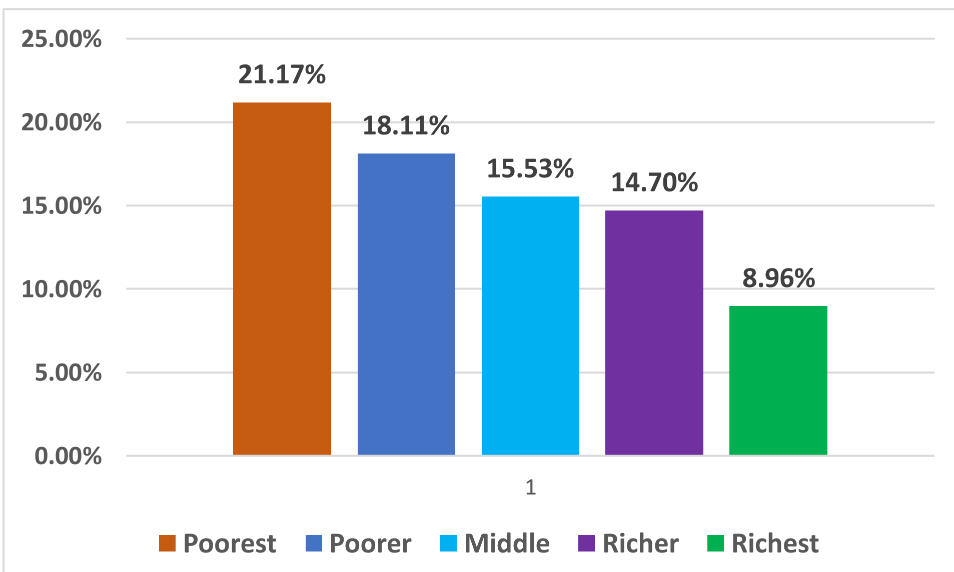

Figure 4.3 Percentage of Infant mortality in different type wealth index of household

Figure 4.3 represent that the most percentages of infant mortality occurs in the poorest households and the least percentage of infant mortality occur in the richest households. Hence the impact of household's wealth index influences in reducing infant mortality.

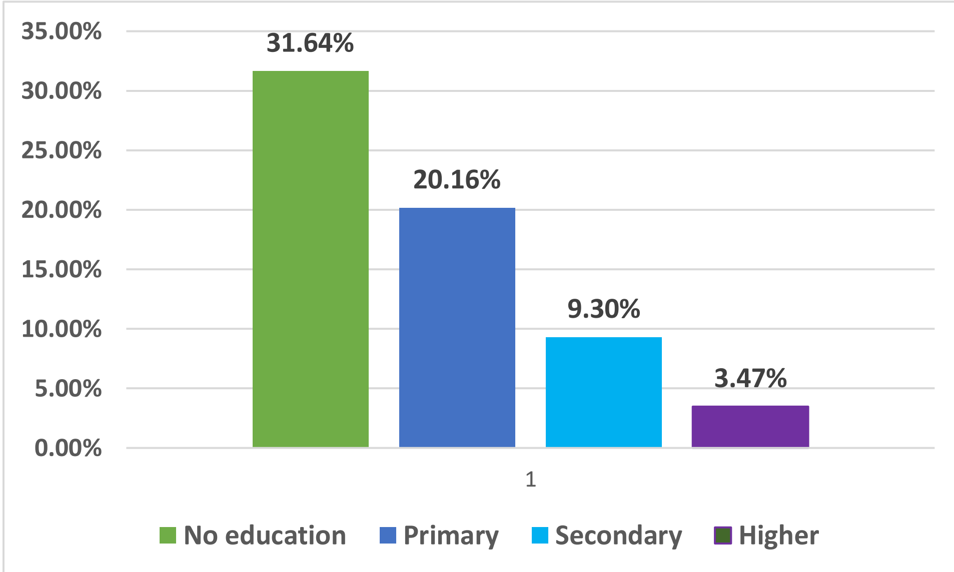

Figure 4.4 Percentage of Infant mortality in different education level of mother

Figure 4.4 represent that the most percentages of infant mortality occurs to the mothers who have less than secondary level of education and the least percentage of infant mortality occur to the mothers who have higher level of education. Hence the impact of mother's education in reducing infant mortality is very high.

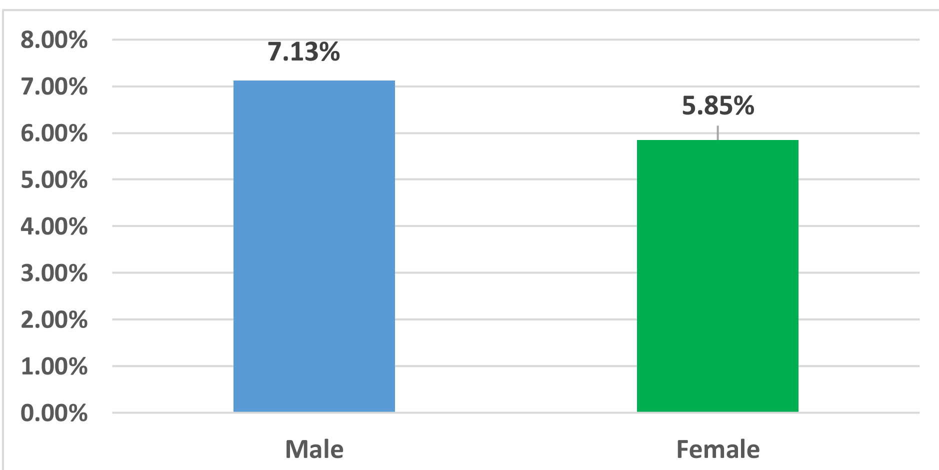

Figure 4.5 Percentage of Infant mortality in different sex

Figure 4.5 represent that the most percentages of infant mortal children are males and the least percentage of infant mortality occur to the female children. Hence the influence of sex in infant mortality is not same.

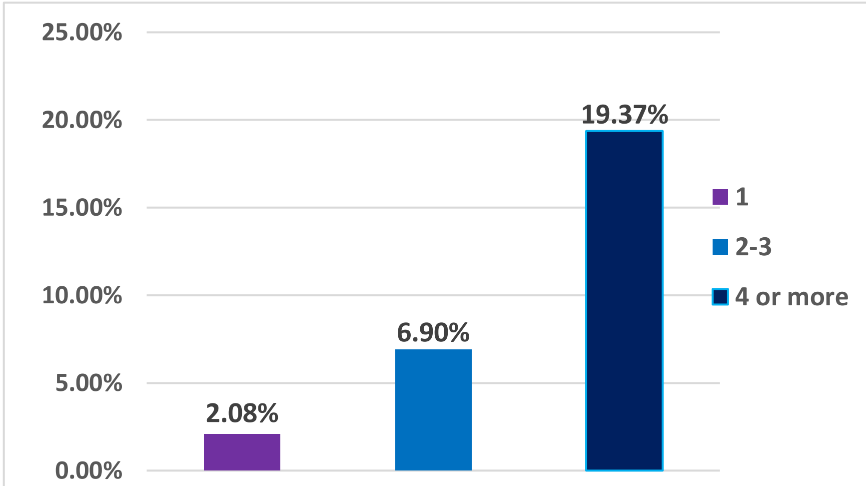

Figure 4.6 Percentage of Infant mortality in different birth order of children

Figure 4.6 represent that the most percentages of infant mortal children are who born in birth order 4 or more and percentage of infant mortality occur to the children who born in first birth order. Hence the influence of birth order of child in infant mortality is not same.

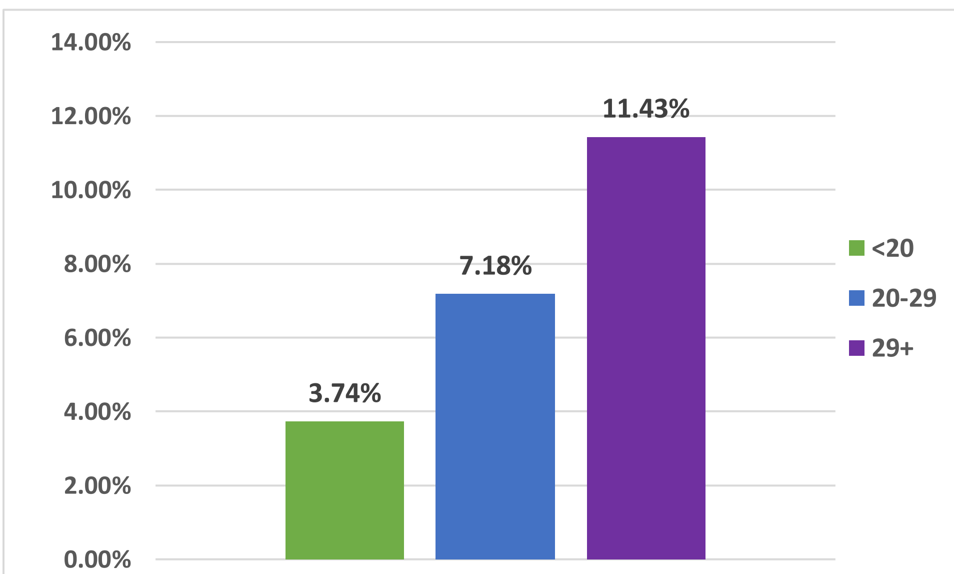

Figure 4.7 Percentage of Infant mortality at different mother age

Figure 4.7 represent that the most percentages of infant mortal children's mothers are aged 30 years or more and percentage of infant mortality occur to the children whose mothers are aged less than 20 years. Hence the influence mother's age at birth in infant mortality is not same.

Table 4.1 Percentage distribution of infant mortality in Bangladesh, 2017-2018

Covariates		Frequency	Infant Death	Percentages (%)	Chi-Square Test
Division	Barisal	5293	324	6.12%	$\chi^2 = 35.15$ $p < 0.001$
	Chittagong	7421	395	5.32%	
	Dhaka	6561	416	6.34%	
	Khulna	5484	331	6.04%	
	Mymensingh	5343	389	7.28%	
	Rajshahi	5607	397	7.08%	
	Rangpur	5911	398	6.73%	
	Sylhet	6208	461	7.43%	
	Total	47828	3111		
Place of residence	Urban	7374	946	12.83%	$\chi^2 = 21.69$ $p < 0.001$
	Rural	12753	2165	16.98%	
	Total	20127	3111		
Highest educational level	No education	3202	1013	31.64%	$\chi^2 = 258.21$ $p < 0.001$
	Primary	6340	1278	20.16%	
	Secondary	7764	722	9.30%	
	Higher	2821	98	3.47%	
	Total	20127	3111		
Wealth index combined	Poorest	3826	810	21.17%	$\chi^2 = 89.16$ $p < 0.001$
	Poorer	3833	694	18.11%	
	Middle	3883	603	15.53%	
	Richer	4088	601	14.70%	
	Richest	4497	403	8.96%	
	Total	20127	3111		
Sex of child	Male	24587	1752	55.85%	$\chi^2 = 28.32$ $p < 0.001$
	Female	23241	1359	44.15%	
	Total	47828	3111		
Birth order of child	1	18134	373	11.99%	$\chi^2 = 1790.2$ $p < 0.001$
	2-3	21866	1437	46.19%	
	4 or more	7828	1301	41.82%	
	Total	47828	3111		
Age of mother at birth	<20	18177	679	3.74%	$\chi^2 = 484.01$ $p < 0.001$
	20-29	22606	1624	7.18%	
	30 or more	5737	656	11.43%	
	Total	46520	2959		

4.2 Differences in infant mortality based on their backgrounds

During the 2017-2018 year, the infant mortality in urban regions was consistently lower than in rural areas. Infant death in 12.83% of metropolitan regions and 16.98% percent of rural locations.

There were significant differences in the infant death between divisions. During 2017-2018, infant death in Chattogram division had the lowest rate while in Sylhet division had the highest rate of infant mortality (Table 1).

During the period of 2017 to 2018, the frequency of infant death was higher 31.64% with no education than lower 3.47% with a higher education.

During the period of 2017 to 2018, infant death in poorest wealth index had substantially higher rate than their counterparts in the richest wealth index (Table 1).

The greater the chi-square value, the more probable it is to be significant, and hence to reject the null hypothesis and determine that the variables are connected. The null hypothesis that all variables are independent of one another is rejected since the p-value in the test result is smaller than the typical alpha value. To put it another way, the conclusion is important: the data shows that the response variable and explanatory variables are connected (Table 1).

Table 4.2. Odds ratio of different covariates with confidence interval

	Covariates	**Odds ratio [confidence interval 95%]**
Division	Barisal	[Reference]
	Chittagong	0.83 [0.71, 0.97]*
	Dhaka	1.14 [0.97, 1.34]
	Khulna	1.20 [1.01, 1.42]*
	Mymensingh	1.15 [0.98, 1.35].
	Rajshahi	1.33 [1.14, 1.57]***
	Rangpur	1.17 [1.00, 1.37].
	Sylhet	1.05 [0.90, 1.23]
Type of place of residence	Urban	[Reference]
	Rural	1.02 [0.93, 1.12]
Mothers educational level	No education	[Reference]
	Primary	0.94 [0.86, 1.03]
	Secondary	0.88 [0.79, 0.98]*
	Higher	0.67 [0.53, 0.84]***
Wealth index of family	Poorest	[Reference]
	Poorer	0.93 [0.83, 1.04]
	Middle	0.93 [0.83, 1.05]
	Richer	0.98 [0.87, 1.11]
	Richest	0.83 [0.72, 0.97]*
Sex of child	Male	1.31 [1.21, 1.41]***
	Female	[Reference]
Mothers age at birth	>19	[Reference]
	20-29	0.88 [0.78, 0.98]*
	29+	0.82 [0.71, 0.95]**

Birth order of child	1	[Reference]
	2-3	3.52 [3.09, 4.01]***
	4 or more	10.26 [8.77, 12.02]***

4.3 Covariates of infant mortality in Bangladesh

Binary logistic regression was performed to assess the impact of several factors on the likelihood that respondents would infant mortality. The model contained seven independent variables (Division, Type of place of residence, Mother education level, Wealth index of family, Sex of child, Mothers age at birth, and Birth order of child). We consider the first value as a REF variable in the Odds ratio table.

Infant in the division, Chittagong, was found to have 17% lower odds of mortality than their counterparts living in the Barisal division. Infant in the Khulna and Rajshahi division was found to have a 19% and 33% higher risk of mortality respectively than their counterparts living in the Barisal division, which is significant. Infant in the Mymensingh division was found to have 15% higher odds of mortality than their counterparts living in the Barisal division.

In the type of residence, infants living in rural areas and urban areas have approximately similar infant mortality rates.

At the mother education level, we were found to have a 12% lower risk for secondary education and a 33% lower risk of higher education than no education respectively.

In sex of child, 31% higher risk for infant mortality to occur for the male child than the female child.

In the wealth index of a family, the poorest wealth index was linked to a greater likelihood of infant mortality. The richest families had a 17% less chance of infant mortality than their counterparts from the poorest wealth index.

In mother age at birth, 20-29 aged women in 13% lower risk of mortality than their counterparts in <19 age group and 29+ age group women in 18% lower risk of mortality than their counterparts in <19 age group.

Children born in the second or third birth order had three times increased risk of infant death counterparts corresponding to children born in the first birth order, while children born in the fourth or higher birth order had ten times increased risk of infant mortality counterparts corresponding to children born in the first birth order.

4.4 Random forest regression model

The mean of squared residuals of the fitted random forest model is 0.05741279 and 4.16% variance explained indicate how well the model fits the data. Residuals are a difference between prediction and the actual value.

To understand what the model has learnt, the importance = TRUE function is used as in the analysis code. Random forest regression in R provides two outputs: Decrease in mean square error (MSE) and node purity.

Prediction error described as MSE is based on permuting out-of-bag sections of the data per individual tree and predictor, and the errors are then averaged. In the regression context, Node purity is the total decrease in residual sum of squares when splitting on a variable averaged over all trees (i.e. how well a predictor decreases variance).

MSE is a more reliable measure of variable importance. If the two importance metrics show different results, listen to MSE.

Table 4.3: Mean square error (MSE) and node purity of the fitted random forest model

Covariate	**% MSE**	**Node purity**
Division	66.40597	102.09855
Type of place of residence	51.71039	42.94050
Wealth index combined	97.76023	88.91583
Mothers educational level	88.60636	64.45029
Age of mother at birth	76.15346	50.25974
Sex of child	19.96449	26.29502
Birth order of child	90.93519	71.19958

The built-in varImpPlot() function is used to visualize the results. Here, both importance measures we are combined into one plot emphasizing MSE results.

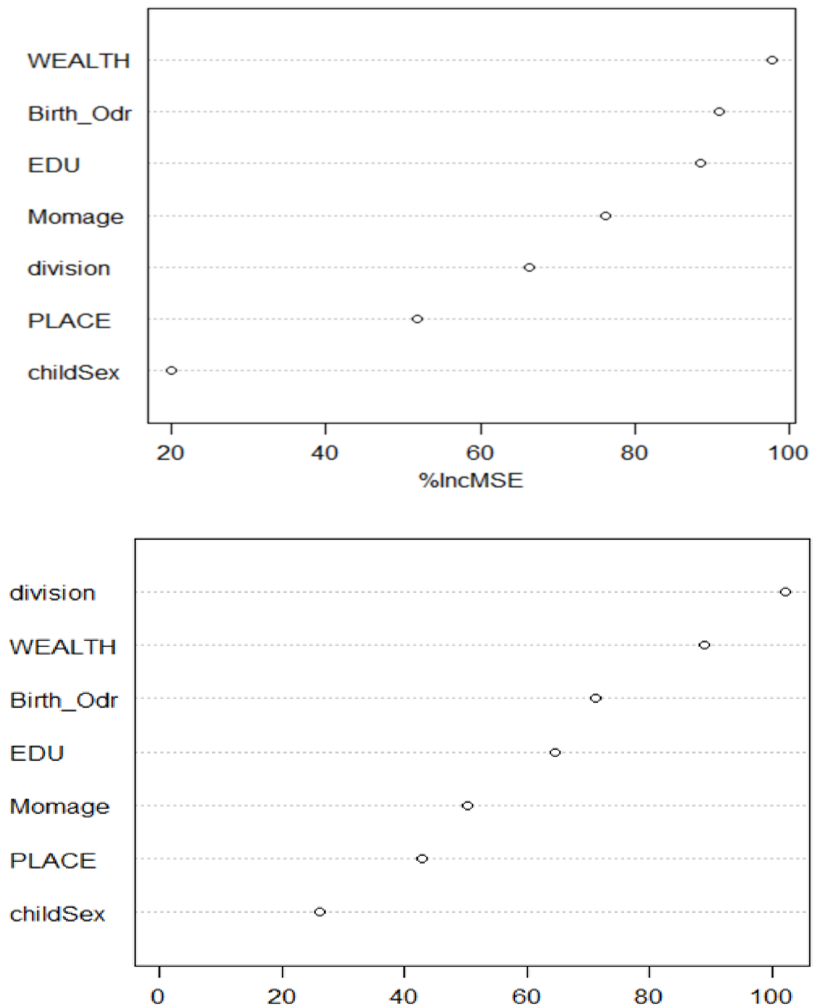

Figure 4.8 Mean square error (MSE) and node purity

From the fitted random forest model it is observed that the most important social-economic factor influencing the infant mortality is the wealth status of the household of the child and the most important birth characteristic influencing the infant mortality is birth order of the children. Whereas, the least important factor influencing the infant mortality is the sex of child.

4.5 Binary logistic regression and random forest regression comparison

Logistic regression is extensively utilized to solve industry-scale issues since it is simple to implement and provides probabilities for each outcome rather than discrete outputs. The logistic regression approach is resistant to minor data noise and is unaffected by minor incidences of multicollinearity. We may evaluate using a variety of measures, but let's focus on True Positive Recall and F1 Score (Harmonic Mean for Recall and Precision)

Random Forest Classifier is a more accuracy-focused algorithm that works best when employed with a good fit; otherwise, it soon becomes overfat. Individual RFC decision trees with random selection can capture more complicated feature patterns and deliver the greatest accuracy. RFC may also use Feature Importance and tree graphs to better understand how much each feature contributes to class prediction.

In this study, a binary logistic regression was fit to fit the BDHS dataset to see the impact of biodemographic and sociodemographic factors on infant mortality. The AIC value of the model was 20290.

When fitting a random forest regression model to the BDHS dataset to see the impact of biodemographic and sociodemographic factors on infant mortality. The Mean of squared residuals was 0.05741279 and the percentage of variance explained by the model was only 4.16. Which represents the lack of fit of the model. To improve the model the number of trees was set to 500 and mtry = 6, but the result did not improve. The AIC value was calculated with MSR and the value was -93034.39.

Comparing both the AIC value we conclude that, the Binary Logistic Regression model fits the BDHS dataset better than the Random Forest Regression model.

Overall, when it comes to categorical data, declaring Random Forest Classifier performs better than numeric and logistic regression is a little ambiguous. If the dataset contains more categorical data and outliers, the Random Forest Classifier should be used.

When the number of noise factors is fewer than or equal to the number of explanatory variables, logistic regression performs better, and the random forest has a greater true and false positive rate as the number of explanatory variables in a dataset grows.

Chapter 5 Conclusion

Bangladesh has achieved significant progress in lowering under-five mortality and inequality over the previous two decades, but the socio-economic imbalance in infant mortality prevents further improvement. The 2017-2018 BDHS dataset considered in this study indicated that birth rank and mother's age at birth and mother's education level and household income index were common factors associated with infant mortality. These findings underscore the need for governments and non-governmental organizations to increase women's education levels to at least the secondary level and take the required efforts to prevent early parenthood. For the benefit of humanity, income disparities must be decreased, and capitalism must be eradicated from the earth. Furthermore, birth interval, child sex, and area of residence are involved in infant mortality. Infant mortality rates in Bangladesh have been shown to differ across districts. The study's findings might assist in establishing a framework for future health plans and policies designing and implementing effective socio-economic measures to improve child survival. The government of Bangladesh and other stakeholders might utilize our findings to assist increase initiatives to minimize infant mortality in Bangladesh.

This study is based on nationally representative household surveys that reflect every locality in Bangladesh. Data were pooled together to create large sample sizes of deaths reported within five years preceding the surveys. Analyses were restricted to births within five years of each of the surveys to reduce recall bias by mothers interviewed and minimize bias that may have arisen from changes in household characteristics. Newborns' dates of birth and death given by mothers may have been misreported, particularly those that had occurred a few months or years before the survey. Causes of death and children's medical conditions were unknown in the study.

References

1. National Institute of Population Research and Training - NIPORT, Ministry of Health and Family Welfare, and ICF. 2020. Bangladesh Demographic and Health Survey 2017-18. Dhaka, Bangladesh: NIPORT/ICF. Available at https://www.dhsprogram.com/pubs/pdf/FR344/FR344.pdf.
2. Abdul Hamid Chowdhury, Prof. Dr. S.M. Shafiqul Islam, Prof. Dr. Abdul Karim, D. (2013). Covariates of Neonatal and Post-Neonatal Mortality in Bangladesh. *Global Journal Of Human-Social Science Research,* Retrieved from https://socialscienceresearch.org/index.php/GJHSS/article/view/739
3. Abir T, Agho KE, Page AN, et al Risk factors for under-5 mortality: evidence from Bangladesh Demographic and Health Survey, 2004–2011BMJ Open 2015; 5:e006722. doe: 10.1136/bmjopen-2014-006722
4. Kabir, R., Farag, M., Lim, H.J. et al. Socio-demographic and environmental risk factors associated with multiple under-five child loss among mothers in Bangladesh. BMC Pediatr 21, 576 (2021). https://doi.org/10.1186/s12887-021-03034-y
5. Khan MA, Khan N, Rahman O, Mustagir G, Hossain K, Islam R, et al. Trends and projections of under-5 mortality in Bangladesh including the effects of maternal high-risk fertility behaviours and use of healthcare services. PLoS One. 2021; 16(2):e0246210.
6. Maniruzzaman M, Suri HS, Kumar N, Abedin MM, Rahman MJ, El-Baz A, Bhoot M, Teji JS, Suri JS. Risk factors of neonatal mortality and child mortality in Bangladesh. J Glob Health. 2018 Jun; 8(1):010417. doi: 10.7189/jogh.08.010421. PMID: 29740501; PMCID: PMC5928324.
7. Khan, J.R., Awan, N. A comprehensive analysis on child mortality and its determinants in Bangladesh using frailty models. *Arch Public Health* 75, 58 (2017). https://doi.org/10.1186/s13690-017-0224-6
8. M Hossain, M Islam. Socio-economic Variables Affecting Infants and Children Mortality in Bangladesh. The Internet Journal of Health. 2008 Volume 9 Number 2.
9. Rahman MM, Ara T, Mahmud S, et AL Revisit the correlates of infant mortality in Bangladesh: findings from two nationwide cross-sectional studies. BMJ Open 2021; 11:e045506. doi: 10.1136/bmjopen-2020-045506
10. Akter T, Hoque D, Chowdhury ME, Rahman EK, Russell M, Arifeen SE. Is there any association between parental education and child mortality? A study in a rural area of Bangladesh. Public Health. 2015; 129:1602–9. doi: 10.1016/j.puhe.2015.08.004.
11. ABDUL HAMID CHOWDHURY, PROF. DR. S.M. SHAFIQUL ISLAM, PROF. DR. ABDUL KARIM, Dr. Covariates of Neonatal and Post-Neonatal Mortality in Bangladesh. Global Journal of Human-Social Science Research, [S.l.], aug. 2013. ISSN 2249-460X. Available at: <https://socialscienceresearch.org/index.php/GJHSS/article/view/739>.
12. Chao F., et al., National and regional under-5 mortality rate by economic status for low-income and middle-income countries: a systematic assessment. The Lancet Global Health, 2018. 6(5): p. e535–e547.

https://doi.org/10.1016/S2214-109X (18)30059-7 PMID: 29653627

13. IGME. UN Inter-agency Group for Child Mortality Estimation. 2018; Available from: https://childmortality.org/data/UNICEF
14. UN, The Millennium Development Goals Report. 2015, The United Nations: New York, USA.
15. You D., et al., Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. 2015. 386(10010): p. 2275–2286.
16. Ezeh O.K., et al., Risk factors for post neonatal, infant, child and under-5 mortality in Nigeria: a pooled cross-sectional analysis. BMJ Open, 2015. 5(3). https://doi.org/10.1136/bmjopen-2014-006779 PMID: 25818271
17. Agho K.E., et al., Factors associated with under-5 mortality in three disadvantaged East African districts. International Health, 2020.
18. Van Malderen C., et al., Socioeconomic factors contributing to under-five mortality in sub-Saharan Africa: a decomposition analysis. BMC Public Health, 2019. 19(1): p. 760. https://doi.org/10.1186/s12889-019-7111-8 PMID: 31200681
19. Hossain M.M., Mani K.K., and Islam M.R., Prevalence and determinants of the gender differentials risk factors of child deaths in Bangladesh: evidence from the Bangladesh demographic and health survey, 2011. PLoS neglected tropical diseases, 2015. 9(3): p. e0003616. https://doi.org/10.1371/journal.pntd. 0003616 PMID: 25747178
20. Ghimire P.R., et al., Under-Five Mortality and Associated Factors: Evidence from the Nepal Demographic and Health Survey (2001–2016). International journal of environmental research and public health, 2019. 16(7): p. 1241.
21. Aheto J.M.K., Predictive model and determinants of under-five child mortality: evidence from the 2014 Ghana demographic and health survey. BMC Public Health, 2019. 19(1): p. 64. https://doi.org/10.1186/ s12889-019-6390-4 PMID: 30642313