MODELLING CHILD MORTALITY VIA DISCRIMINANT ANALYSIS AND LOGISTIC REGRESSION

Arumapperuma Kankanamge M. Dinithi P. Kande Arachchi

(168835H)

Master of Science in Business Statistics

Department of Mathematics

Faculty of Engineering

University of Moratuwa

Sri Lanka

February 2021

MODELLING CHILD MORTALITY VIA DISCRIMINANT ANALYSIS AND LOGISTIC REGRESSION

Arumapperuma Kankanamge M. Dinithi P. Kande Arachchi

(168835H)

Dissertation submitted in partial fulfillment of the requirements for the Master of Science in Business Statistics

Department of Mathematics

Faculty of Engineering

University of Moratuwa

Sri Lanka

February 2021

ABSTRACT

Prevalence of deaths of children has particularly become a global concern in strategic decision making in the field of health sector. In Sri Lanka, the risk of deaths at childhood period was higher during the past few decades. Many studies have concerned about the child mortality in various perspectives. The purpose of this study is to find the significant factors on under-five mortality and to recommend a most suitable statistical model to predict the child mortality, under aged 0-5 years of age. The secondary data was collected from the demographic and health survey (2016) conducted by the Department of Census and Statistics (DCS), Sri Lanka. Two types of statistical models: linear discriminant model and binary logistic model are statistically evaluated. Two models were evaluated with classification accuracy, ROC curve, sensitivity/ specificity and sample size variations. Both methods found that, gender of child, marital status, mother's literacy, status of antenatal care, delivery type, pregnancy duration and decision-making ability are significantly influential variables ($p < 0.05$) on the status of child mortality. According to the classification results produced by models, discriminant model correctly classified the 89.6% of grouped cases and binary logistic regression model correctly classified the 94.6% of grouped cases irrespective of the status of child mortality. With respect to the all seven indicators, it was found that binary logistic regression model was more efficient and more effective than linear discriminant model. The inferences derived can be effectively used for strategic decision making in the health sector for reducing the child mortality in the future.

Keywords: AUC, Binary Logistic Regression, Child Mortality, DHS Survey, Discriminant Analysis, Misclassification, ROC, Sample Size, SDG Goal

ACKNOWLEDGEMENT

Apart from the efforts of me, the success of this research depends largely on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this project.

First and foremost, I would like to thank to my supervisor Senior Prof. T. S. G. Peiris, Senior Professor in Applied Statistics and former Head of Department of Mathematics, Faculty of Engineering, University of Moratuwa and the Course Coordinator of M. Sc. / Post Graduate Diploma in Business Statistics (2016/2017) batch for the valuable guidance and advice. He inspired me greatly to work in this project. His willingness to motivate me contributed tremendously to this project.

Then, I wish to offer my greatest appreciations towards the lecturers and members in the staff of the Department of Mathematics, University of Moratuwa, for the admirable support, during the degree program.

I would like to express my gratitude towards the staff members of Statistical Division, Department of Census and Statistics, Sri Lanka who supported me with their kind co-operation and encouragement which helped me in completion of this project. I would like to extend my sincere thanks to all of them.

My Special thanks goes to Dr. K.J.L.Kuray and Dr.Lasantha Somarathna, who helped me at the last moment for the achievement of the successful ending of this course work.

Finally, an honorable mention goes to my family for their understandings and supports on me in completing this dissertation.

DECLARATION OF THE CANDIDATE

"I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text".

Also, I hereby grant to University of Moratuwa, the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works.

....   06/02/2021

Signature   Date

A.K.M.D.P. KandeArachchi

Reg. No: 168835H

DECLARATION OF THE SUPERVISOR

The Dissertation entitled ‘Modelling Child Mortality via Discriminant Analysis and Logistic Regression’ has been carried out by the candidate, A.K.M.D.P. KandeArachchi (Reg. No: 168835H) for the submission of Master’s Degree under my Supervision.

…………………………..   06/02/2021

Senior Prof. T.S.G. Peiris   Date

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

Abbreviation	Description
ACCR	Apparent Correct Classification Rate
AER	Apparent Error Rate
AIDS	Acquired Immunodeficiency Syndrome
ANOVA	Analysis of Variance
AUC	Area Under Curve
BCG	Bacille Calmette Guerin
CI	Confidence Interval
CM	Childhood Mortality
DA	Discriminant Analysis
DCS	Department of Census and Statistics
DHS	Demographic Health Survey
DTP	Diphtheria and Tetanus Toxoids and Pertussis
FHB	Family Health Bureau
FLD	Fisher's Linear Discriminant
G.C.E. (O/L)	General Certificate of Education (Ordinary Level)
GDP	Gross Domestic Product
HIV	Human Immunodeficiency Virus
IGME	Inter-Agency Group for Child Mortality Estimation
IMR	Infant Mortality Rate
LDA	Linear Discriminant Analysis
LDF	Linear Discriminant Function
LR	Logistic Regression
MANOVA	Multivariate Analysis of Variance
MCH	Maternal and Child Health
MN	Million
MLE	Maximum Likelihood Estimate
MSD	Multivariate Statistical Distance
QDF	Quadric Discriminant Analysis
ROC	Receiver Operating Characteristic
SCFA	Short-Chain Fatty Acids
SDG	Sustainable Development Goal
SL	Sri Lanka
SS	Sums of Square
UN	United Nations
UNICEF	United Nations Children's Fund
VIF	Variance Inflation Factor
WHO	World Health Organization

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

Around the world, millions of new-born and children under five years unacceptably die, most probably by the preventable and treatable causes yearly. (United Nations, 2017). Conventionally, it is measured as a death under the particular age, divided by the number of live births per year. In Sri Lanka, around 328,000 registered live births are happened every year (Family Health Bureau, 2016).

In relation to the recent health statistics in Sri Lanka, according to the UNICEF estimates, 1509 neonatal deaths, 2148 infant deaths and 2521 under five deaths have been reported in 2018 by giving a neonatal mortality rate of 4.49 per 1,000 live births and an under-five mortality rate of 7.44 per 1,000 live births. Further it is evidenced that under-five mortality rate has the greater declining trend from 97.85% to 7.44%, by exposing the rapid change in development of health sector in Sri Lanka over the period of 1960-2018. Globally, it was highlighted under-five child mortality rate as 41 per 1000 live births in 2016.

In Sri Lanka, most of the child deaths are occurred within the first month of the child birth and the risk is pertaining until the first birth day of the child, during the infancy period (SLDHS, 2016). As the leading causes, in 2016, approximately 18% of children aged 0-5 were diagnosed with preterm birth disorders, congenital malformations and deformations. More than 10% of children were infected by the second level of leading causes indicating pneumonia and other bacterial diseases in Sri Lanka.

Focusing these study issues, there are abundant evidences available in literature. This is a one of the simplistic efforts to try out the modelling child mortality under aged 0-5 by modelling the identified determinants via discriminant analysis and the binary logistic regression analysis in Sri Lanka based on the DHS survey data, year 2016. From this study analysis, it will be provided the suggestions in improvement of the infant and child survival for constructing community based public policies.

1.2 Child Mortality in Sri Lanka

1.2.1 Movements in Child Mortality Rates, Aged 0-5

Since the Second-World-War, Sri Lanka has experienced a long-run downward trend in mortality rates of children, which are attributed to the strong policy implications in the health sector, towards eradicating the study problem. For the studying purpose, deaths of children are categorized into five different pre-defined age-specific sectors as follows:

- **Neonatal Mortality**: Deaths, within first month from child birth
- **Post-Neonatal Mortality**: Deaths, between completed first month and first year of a child (*Difference between infant mortality and neonatal mortality*)
- **Infant Mortality**: Deaths, before the first 12 months from the child birth
- **Child Mortality**: Deaths, between the completed first year and fifth year of a child
- **Under-Five Child Mortality**: Deaths, from the child birth until the fifth year of a child (*Summation of infant mortality and child mortality*)

Table 1.1: Age Specific Childhood Mortality Rates in SL, 2013-2018

Time Period	2013	2014	2015	2016	2017	2018
Neonatal Mortality Rate:0-1 m	5.83	5.56	5.26	4.97	4.71	4.49
Post-Neonatal Mortality: 1-12m	2.74	2.45	2.23	2.08	1.96	1.87
Infant Mortality Rate: 1q0	8.57	8.01	7.49	7.05	6.67	6.36
Child Mortality Rate: 4q1	1.44	1.35	1.27	1.20	1.14	1.08
Under-Five Mortality Rate: 5q0	10.01	9.36	8.76	8.25	7.81	7.44

Source: UN IGME (2018)

To the extent, Table 1.1 depicts age specific child mortality rates within the duration of 2013-2018 in Sri Lanka based on the UN estimates. Comparatively, child mortality rate was significantly lower with respect to other four categories and it was ranged from 1.44 per 1000 live births to 1.08 per 1000 live births over the study period and under-five mortality has the significant improvement in declining trend. The highest declination indicated the under-five mortality rate, 2.57 per 1000 live births and the lowest by child mortality rate, 0.36 per 1000 live births.

Infant mortality rates and under-five mortality rates were generally constituted relatively higher values within the study period in Sri Lanka. Infant mortality rate was attributed with 7.08 per 1000 live births of female and 8.93 per 1000 live births of male to the average mortality rate: 8.02 per 1000 live birth in 2019, which was beyond the pattern of the previous declining trend reported by the UN estimates.

In 2013, post-neonatal mortality rate was estimated as 2.74 per 1000 live births and declined into 1.87 per 1000 live births in 2018. Continuous maintenance of lower rates of child deaths, which were reported within 28 days from the child birth is a great achievement of the health sector in Sri Lanka.

1.2.2 Socio-Demographic Characteristics of Child Mortality

Child mortality rates are the set of indicators for measuring level of social, economic and environmental situations in the society. Table 1.2 and Table 1.3 demonstrate the selected socio-demographic factors of the study objectives, indicating the information related to the DHS survey, preceding the period of 10 years from the year 2016 including the gender of child, birth order and education of mother under the categories of children, aged 0-1 and children, aged 1-5 with the under-five mortality rates.

Table 1.2: Under-Five Mortality in SL, by Gender, 2016

		Under Five Mortality Rate	Percentage of Contribution, to the Total Deaths (%)		Total No of Deaths
			Infants, aged 0-1	Children, aged 1-5	
Gender	Male	14	75	25	469
	Female	10	74	26	314

* Under Five Mortality Rate (Deaths per 1000 live Births), 10 yr period preceding the survey, 2016

Source: DHS (2016)

Table 1.2 shows the percentage distribution of under-five child deaths by gender. According to the Table 1.2, female contribution is in the lower level, 40.1 % (314) than that of the male contribution, 59.9% (469) of the overall population of children aged 0-5. Male child (14) represents the higher level of mortality rate compared to the female child (10) in Sri Lanka in the study period. Higher level of contribution indicates the lower level of survival status. The results of percentage distribution show the majority of deaths are belonged to the population, aged 0-1 means the increased risk of infants due to clinical and health issues of mother and child from the perinatal period.

Table 1.3: Under-Five Mortality in SL, by Birth Order & Mother's Education, 2016

	Under Five Mortality Rate	Percentage of Contribution, to the Total Deaths (%)				Total No of Deaths
		Infants, aged 0-1		Children, aged 1-5		
		Male	Female	Male	Female	
Birth Order						
1	12	47	26	17	10	363
2-3	11	43	33	13	11	337
4-6	21	43	35	13	9	79
Education of Mother						
Passed Grade 1-5	14	37	32	17	14	172
Passed Grade 6-10	12	46	28	16	10	375
Passed G.C.E.(O/L) or Equivivelent	12	48	27	14	11	124
Passed G.C.E.(A/L) or Equivivelent	11	48	39	8	5	66
Degree and Above	6	56	44	0	0	9

* Under Five Mortality Rate (Deaths per 1000 live Births), 10 yr period preceding the survey, 2016
Source: DHS (2016)

Table 1.3 represents the percentage distribution of under-five child deaths by birth order and education of mother. Considering the under-five child mortality rates, in terms of the birth order, demographic characteristic, it was determined 12 per 1000 live births for the first birth order and respectively it was indicated 11per 1000 live births and 21 per 1000 live births as the birth order increased into 2-3 and 4-6. The mortality risk is higher for male children than that of female children irrespective of the age group. The findings reveal that the lesser fertility preferences can be reduced the death rate of children.

With regard to the levels and differentials of education of mother, the lowest rate of mortality showed by the category of degree and above, 6 per 1000 live births and two different categories with different levels of mortality contributions indicated the similar level of mortality rate, 12 per 1000 live births with respect to the categories of passed grade 6-10 and passed G.C.E (O/L) or equivalent. Further, there is a higher risk of mortality when leaving the school, at the secondary level of education, aged 11-15 compared to the other levels of schooling. The relative mortality risk of the number of people who were obtained the degree or above and the advanced level qualification are comparatively lower which indicate the 1.2% and 8.7% respectively of total population in Sri Lanka. Then the infant mortality is higher than the child mortality, followed by the education of mother. To the extent that the educated women has the greater ability of protecting their children than the woman by no education pursues due to better decision-making skills about the health matters of the mother and child.

1.2.3 Comparative Review of District-Wise Births and Deaths

Measuring the health status of the society is a one of the refined indicators in economic development in developing countries. To extent to which, health of child is a vital component in the health status of the country's population in terms of nutrition, child development, physical and mental stability. Paying the attention to the problems in the population growth, under-five child mortality is being attributed to the adverse health outcomes.

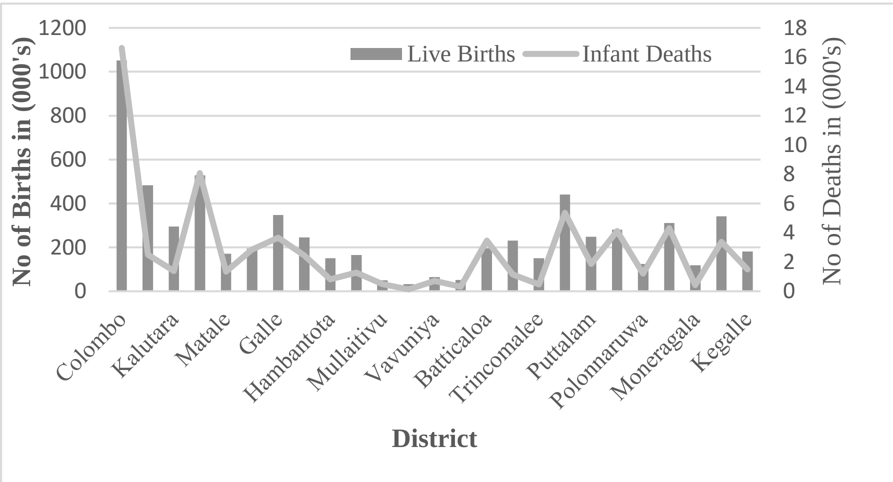


Figure 1.1: District Disparities in Live Births &Infant Deaths, 1997-2014

Source: DCS (2014)

Sri Lanka is currently a developing country, World Bank reported 0.61% in 2019 as the annual population growth rate to the total population 21.9 million according to the census figures. It was growing by 0.42% from 2019 to 2020 whilst 7.4 per 1000 live births was reported in 2018 as the under-five mortality rate.

In 2018, it was reported the 14.8 births per 1000 population, 0.73% as the population growth rate and 8.2 deaths per 1000 live births as the infant mortality rate in Sri Lanka. As the Figure 1.1 displays that relative to the other districts, Colombo has the highest percentage of reporting, live births and infant deaths parallel with the greatest population growth, nearly 2.3 million in 2020 as the commercial capital city followed by Kandy district at central province in Sri Lanka.

The region with the lowest births and deaths were experienced at Mannar district. Analysis depicts, during the last two decades, Colombo and Badulla districts were with greater infant mortality declines, while low level of decline in Kandy and unstable trend in Kurunegala. The lowest percentage of births and deaths occurs in Eastern province, districts including Mannar, Vavuniya, Mullaitivu, Kilinochchi based on overall population.

Except Colombo, total live births have not demonstrated any significant trending pattern from 1997 to 2014. All child deaths were confirmed and reported by hospitals. There are wide district disparities when it is categorized, both live births and infant deaths into urban, estate and rural according to population gap, development issues etc. These highlights are more important in reducing mortality while event of occurring child birth.

1.3 Child Mortality in Global Context

1.3.1 Global Analysis of Childhood Mortality

As United Nations stated that SDG goal 4 targeting to conclude the preventable under-five child deaths by 2030. Since 2012, considering the sustainable development goal regions in the world, under-five mortality rates have shown substantial decrease until 2018 as the Table 1.4 with wide variations of the mortality rates among the selected regions.

Within the study period, North America & Europe and Australia & New Zealand were stable in mortality rates by maintaining approximate value around 5 per 1000 live births and the overall declination was higher. It was highlighted that the initial mortality rate was lower in Australia & New Zealand and higher in least developed countries and that pattern of variation in mortality continued until 2012. In relation to the trends and patterns of the overall distribution, the difference between least developed and developed countries found to be reduced in the world.

Table 1.4: Under Five Mortality Rates by Selected SDG Regions, 2012-2018

Region Name	2012	2013	2014	2015	2016	2017	2018
Sri Lanka	10.61	9.99	9.35	8.75	8.23	7.80	7.44
Least developed countries	80.60	77.30	74.25	71.46	68.80	66.40	64.23
Northern America and Europe	6.52	6.37	6.21	6.04	5.88	5.71	5.55
Southern Asia	54.98	52.33	49.80	47.39	45.11	42.93	40.97
Western Asia and Northern Africa	30.43	29.55	28.74	27.97	27.27	26.57	25.92
Australia and New Zealand	4.61	4.44	4.31	4.20	4.13	4.07	4.01

Source: UN IGME (2018)

During the period of 2012-2018, least developed countries, Southern Asia and Western Asia & Northern Africa pursued higher rates in mortality than that of other 2 regions and the overall declination was higher too. Sri Lanka, under-five mortality rate fell gradually from 10.61 per 1000 live births, in 2012 to 7.44 per 1000 live births, in 2018 and the percentage change was -5.13% compared to year 2017. Changes in health policies in Sri Lanka were significantly influenced on the decline in the mortality rates and continuously trying to reach the levels on par with industrialized countries in the world.

1.3.2 Leading Causes of Child Mortality Worldwide

Causes of deaths of children are varied geographically depending on the existing determinants in the area as stated by world health organization. Even though the child mortality is a global issue, actions for the precautions need to be taken within the country level. Table 1.5 illustrates the global distribution of child deaths by the causes, year 2016. UNICEF (2017) reported that from the total distribution of medical causes of child deaths, the estimate of less than 10% is accounted with least affected diseases on children as Malaria, Tetanus, Measles and AIDS.

Then, it is noted that Preterm birth compilations and Pneumonia are the major reasons of global childhood mortality, which contributes 18% and 16% of deaths respectively. The other causes of child mortality indicate Intrapartum related events, Sepsis or Meningitis, Congenital, Diarrhea and Injury which are directly or indirectly supported to increase the level of deaths throughout the world. 10.7% of deaths are constituted the other complications of health of the mother and child throughout the study period.

Table 1.5: Global Distribution of Child Deaths by Causes, 2016

	Neonatal deaths (%)	**Children Aged, 0-5 (%)**
Preterm birth complications	16.0	2.0
Pneumonia	3.0	13.0
Other	3.0	7.7
Intrapartum-related Events	11.0	1.0
Sepsis or meningitis	7.0	2.0
Congenital	5.0	4.0
Diarrhea	0.3	8.0
Injury	1.0	6.0

Source: UNICEF (2017)

Furthermore, according to the statistics of F.H.B., 2013 in Sri Lanka, 43% of infant deaths were occurred due to the congenital deformities, 11% asphyxia and 26% prematurity. In 2016, infant death rate has been increased from 43% to 54.5%.

1.4 Life Expectancy at Birth

Life expectancy at birth is a key indicator, which is measuring the health status of the community in the region and globe. Worldwide perspectives, Japan people advantage in living longer than other countries. United States people are in the average level and African people have the shortest natural life. It was highlighted that; men live shorter period of time than women. Total life expectancy, both men and women together are in the middle level, longer than men & shorter than women. Table 1.6 indicates the comparison of life expectancy at birth of men and women, within the period of 2012-2018

Lower life expectancy may result higher child mortality rate and that leads the lower population growth. Conversely, the gender disparity in life expectancy also has an adverse impact on child survival. It was worst that, Central African Republic reported comparatively lower life span than that of all other countries in the world and during the study period, average life expectancy, changed from 48 years to 52 years, which was slightly negligible.

Table 1.6: Life Expectancy at Birth by Highlighted Counties, 2012-2018

Country Name	**Indicator Name**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**
Japan	Male	79.94	80.21	80.50	80.75	80.98	81.09	81.25
	Female	86.41	86.61	86.83	86.99	87.14	87.26	87.32
	Total	83.10	83.33	83.59	83.79	83.98	84.10	84.21
United States	Male	76.40	76.40	76.50	76.30	76.10	76.10	76.10
	Female	81.20	81.20	81.30	81.20	81.10	81.10	81.10
	Total	78.74	78.74	78.84	78.69	78.54	78.54	78.54
Central African Republic	Male	47.03	47.68	48.34	48.99	49.59	50.15	50.65
	Female	50.24	51.07	51.93	52.79	53.61	54.35	54.99
	Total	48.64	49.37	50.13	50.88	51.59	52.24	52.81
Sri Lanka	Male	72.36	72.54	72.71	72.89	73.06	73.24	73.42
	Female	79.24	79.39	79.54	79.69	79.84	79.98	80.12
	Total	75.80	75.97	76.15	76.32	76.48	76.65	76.81

Source: World Bank (2018)

Sri Lanka, the average life expectancy at birth was male: 73 years and female: 80 years in 2018, life expectancy gap pertaining 7 years and total life expectancy indicating 76 years at birth according to the work bank report. During the study period, 2012-2018, the average change of the life expectancy was insignificant. It was noted that United States, life span of people remained steady state behavior during the study period.

1.5 Importance of Fertility and Mother Care

When the pregnancy is suspended, it is must, mothers receiving the Antenatal care from a skilled healthcare service provider before the completion of 3 months of the pregnancy. Among the Countries in the world, the highest levels of fertility rates are floated in Niger, Somalia, Congo and Mali in 2018 as the word bank reported.

Sri Lanka it was reported the fertility rate as 2.19 per woman in 2018, though there was a continued fertility decline from the past few decades. Maintenance of Birth intervals are common when analyzing the wellbeing of the mother & child and the fertility process too. Women in the country has a strong per capita income, would have less children at her childbearing period and the less developed countries with no or less education leads to have more children in her life time, specially aged 15-49.

Table 1.7: Fertility Rate vs Infant Mortality Rate, 2012-2018

		2012	2013	2014	2015	2016	2017	2018
Korea, Rep.	Fertility	1.30	1.19	1.21	1.24	1.17	1.05	0.98
	Mortality	3.30	3.20	3.10	3.00	2.90	2.80	2.70
Niger	Fertility	7.38	7.31	7.25	7.17	7.09	7.00	6.91
	Mortality	57.20	55.20	53.50	52.00	50.50	49.10	48.00
Sri Lanka	Fertility	2.23	2.22	2.22	2.21	2.21	2.21	2.20
	Mortality	9.10	8.60	8.00	7.50	7.00	6.70	6.40

Source: World Bank & UN IGME (2018)

Preventing the infant deaths, fertility is being handing major role in overall mother and child care process. Table 1.7 illustrates that the exceptionally high rate of mortality in Niger was attributable mostly to unusually high infant mortality rate, in Korea, lower fertility rate leading to exceptionally low mortality rate. Niger's unacceptable high mortality rates were changing 57.20 per 1000 live births to 48 per 1000 live births, while fertility rates were changed 7.38% to 6.91% from 2012 to 2018 by attributing to the poor health status of the community as a whole. Korea, it was constituted lowest infant mortality rates in relation to the lowest fertility rates during the period, which implies the higher human development index. Sri Lanka has an average level of association between fertility rate and infant mortality rate. It was observed that 0.05% decline in fertility rate from 2017 and the average number of children pertaining to a family was two, in 1970 it was reported as four.

Postnatal care is important too, receiving within 2 days after delivery for the protection of new born and mother as well. Guidelines are provided by the healthcare providers about the visits, feeding the child, family planning as according to the mother and the child's health conditions until at least after the delivery of one year, then screening the mother and child for the identifications of disparities.

1.6 Research Problem

In the field of medical science, more resources are available in the literature by making smart choices on study objectives in overcoming many health problems. Most of the studies of child health outcomes are supported the evidences of child malnutrition, child health development, early childhood education etc. In Sri Lanka, focusing the under-five child mortality, it needs much more attention on conducting research studies to the extent to which how it is possible to reach SDG goal 4 in the near future.

Considering the statistical methodologies have been employed in the analysis of existing literature focusing the child mortality Sri Lanka. Many studies were conducted with the application of logistic regression and very less number of studies, obtaining the result outcomes with discriminant analysis and the comparative review outcomes with logistic regression & discriminant analysis aligned with the proposed study objectives.

As described above, child mortality rate is a one of the sensitive societies measurements, which enables to identify the country's health status, distribution of available resources among the citizens, education and economic development, child nutrition etc. in Sri Lanka. Government is responsible in reducing the mortality rates in the prevailing community, when upgrading the country with economic and human development, subsequently increases the child survival.

Targeting to achieve the global challenge of the child survival by reducing infant and child mortality according to the Sri Lankan context, it needs to be identified the socio-economic, demographic and environmental factors of child health outcomes in Sri Lanka first.

1.7 Objectives of the Study

On view of the explanation given above in detailed, the objectives of this study are:

- To develop a suitable statistical model for the child mortality
- To validate the developed model
- To give recommendations and suggestions for national health policy decision makers with regards to child mortality

1.8 Significance of the Study

The inferences expected to derive from the statistical analyses would be beneficial for the followings.

- To reduce child mortality, maintenance of up to-date accurate information repository assist the government to control productive activities towards the global achievements.
- To find the solutions for above mentioned problems and limitations of the current health system in Sri Lanka, which were reflected from the previous studies on the under-five child mortality conducted by applying various statistical approaches.
- To contribute some insights to the existing literature.
- To provide some insight for the health policy makers, authorized people, etc. who are responsible in establishing and development of health system in the country.

1.9 Outline of the Dissertation

This dissertation is structured as follows in eight chapters.

Chapter 1 indicates briefly, background and general introduction to the study related to the theoretical concepts and the practical situation of the focused study, the problem statement and objectives of the study.

Chapter 2 takes a look at the theoretical framework and reviews related literature concerning the study.

Chapter 3 gives an overview of the particular research methodology with brief description of theoretical concepts used to achieve the objectives of this study.

Chapter 4 deals with the analysis of risk factors associated with the study.

Chapter 5 discusses the development of modelling the child mortality using discriminant analysis approach.

Chapter 6 discusses the development of modelling the child mortality using logistic regression analysis approach.

Chapter 7 focuses on the comparative analysis between discriminant analysis and logistic regression.

Chapter 8 presents conclusions and suggestions of the study.

CHAPTER 2

REVIEW OF LITERATURE

This Chapter gives review of definitions of theoretical terms and background for the proposed research issue which is based on previous research findings. This is the core part of this study and the purpose of this literature review is the accomplishment of framework by covering the focused areas of the child mortality with discriminant analysis and logistic regression modelling.

2.1 Child Mortality in Sri Lanka

The term of 'Child Mortality' can be defined as the probability of dying children within the pre-defined age groups: neonatal, post-neonatal, infant, child & under-five and it is indicated as the measure of 1000 live births. In Sri Lanka, under-five child mortality was reported as 21 in 2006, then it was declined into 11 within a decade with constitution of 64% infant deaths and 55% aged 1-5 child deaths as stated by Silva (2015) and SLDHS annual report (2016). Therefore, it is automatically forced by the nature to reduce the neonatal deaths before addressing the issue of infant deaths or under-five deaths.

As a third world country, Sri Lanka in relation to other developing countries in the world had the greater achievements in health sector reforms with increasing the life expectancy as for males and females was 43.9 and 41.6 years respectively in 1945, then 69.5 and 74.2 years for males and females, in 1991 by reducing child and maternal mortality which were 263 per 1000 live births and 165.2 per 10000 live births respectively in 1940's, then 15.9 per 1000 live births and 2.3 per 10000 live births respectively in 1996, while experiencing the reduction of non-communicable deceases with increasing the per capita income of US$ 837 for the greater demand for the expenses of the health care treatment procedures over the past few decades (Samarakoon et al., 2015 and Gunasekara et al., 2000).

Weerasekara et al. (2019) determined 45.6% of neonatal deaths within first 24 hours of child birth, duration of 10-year period from 2008 in Sri Lanka in their research. Prematurity, Congenital malformations and respiratory system syndromes were attributable to the leading causes of neonatal deaths and it is lower than to the global and neighbours, south Asian figures. In a previous research by Rajindrajith et al. (2009) came up with the result of the occurrence of 90.5% of neonatal deaths happen within the first week of child birth, period of 1997-2001 and most of them have been reported at the state sector in Sri Lanka. Therefore, the national guideline for new-born care was introduced by the ministry of health by establishing six neonatal intensive care units island wide Sri Lanka in 2014 (Save the Children, 2017).

Reduction of child mortality has been declared as the target of Millennium development goal 4, constituting 3 criteria: under-five mortality rate, infant mortality rate and the proportion of one-year old children immunised against measles nevertheless the availability of incomplete still birth registration system records in Sri Lanka. It was reported that high immunisation and breastfeeding rates are contributed with high literacy rates of mothers towards the increasing child survival. Consequently, maternal and child health (MCH) targets can be achieved by investments in the sectors of free education and free health care services as suggested by Senanayake et al. (2011).

2.2 Global Child Mortality

Global under-five child deaths have been fallen from 19.6 mn. in 1950, 8.5 mn. in 2000 to 5.4 mn. in 2017 with the contribution of higher rate in post-neonatal mortality than the neonatal mortality, truly with the worst performance of countries in sub Saharan Africa. Average under-five child death rate remained 39 per 1000 live births, neonatal mortality rate, 18 per 1000 live births and the infant death rate, 29 per 1000 live births in 2017 (Sanghera, 2018), down from 90 per 1000 live births, under-five deaths and neonatal mortality rate 36.6 per 1000 live births in 1990 worldwide (Mmusi-Phetoe, 2016).

Geographically the risk of these child deaths differs from country-to-country and within the country specific locations. In 2017, as the highest neonatal death rate was reported in Central African Republic, 123.9 per 1000 live births and the lowest rate from Cuba, 5.1 per 1000 live births as examined by Burstein et al.et al. (2019) in their study. Some countries in the world shows huge imbalances of acquiring healthcare facilities, food & nutrition and sanitation & hygiene among the states such as India, which has been reported the under-five child death rate similar to the global rate (Sanghera, 2018).

Sub Saharan Africa and South Asian regions are responsible in more than 80% of under-five child deaths all over the world, hence most of the developing countries suffered from high level of child mortality. Countries like Ethiopia, Nepal (20 per 1000 live births) has been reduced their child mortality level, however still above the SDG goals as reported by Ghimire, 2019.

Despite of higher per capita income, performance of the child health care system remains in a lower level in United States in relation to other industrialized countries in the world such as Japan, China, Russia, Sweden etc. and the risk of infant deaths was 76% greater within 10-year period, even though the industrialized countries approach to the health sector problems in a substantially different way as stated by Thakrar et al.et al. (2018) and Liu et al.et al. (1992). When Western Europe taking into the account United Kingdom has the highest child mortality (Zylbersztejn et al., 2018) due to prematurity, congenital malformations etc.

Inadequacy of the public health decision making system for information gathering is the most problematic section for the achievement of public health targets by reducing the risk of child mortality in most countries at regional and global level as identified by Morris et.al. (2003) in their research. Public health decisions are associated with SDG goals. Targets under the sustainable development goals can be achieved with socio economic developments under the areas of enhancing health education & strategies, woman empowerment and promoting maternal and child health with special focus on breastfeeding, family planning and sanitation & hygiene (Dendup et al., 2018) by prevalence of child deaths entirely in 2030 (Burstein et al., 2019).

2.3 Common Risk Factors of Child Mortality

Geneti and Deressa (2014), Islam et al. (2013) and Rahman et al. (2019) developed a framework for the identification of determinants of the child mortality by the combination of social factors, demographic factors and environmental factors. Tagoe et al. (2020) and Zewdie and Adjiwanou (2017) conducted their study based on the socio-economic and demographic factors. Ali (2005) presented a paper at a population conference, which was based on socio-economic and environmental factors on child mortality.

Hill (1992) and Becher et al. (2004) in their studies focused on the different characteristics of the measurement of factors for measuring the child mortality. It was revealed the importance of the variety of approaches, quality of data collection, accurate measurements in Identification & Classification of Child mortality and the necessity of well-developed vital registration systems for developing countries.

The combined effect of different determinants could be revealed child health in different angles. In this study objective, determinants of child mortality are categorized into three different sectors: demographic factors, socio-economic factors and health/environmental factors. Several studies have focused on the significant relationship between the under-five mortality rate and the identified determinants of the study objectives. In this section, selected predictors are discussed by categorizing into major divisions of common factors.

2.3.1 Demographic Factors

Gender Patterns of Child

A society with more gender inequality discloses the truth of lesser chances of survival in that particular group of children. Most probably excess of female is the problematic situation to be faced in low- and middle-income countries in the world. This aggregates on the occurrence of cultural issues, distribution of resources and the limited social power as documented by Iqbal et al. (2018). In Vietnam, survival rate of girls are higher than boys which in case of 30%. Despite of having Urban- rural gap in mortality. Mother's education influence more on girl's survival rate by Pham et al. (2011).

Compared to the male child population, girl: infant and child population is victimized more due to intimate partner violence in Indian society. Silverman et al. (2011) emphasized to prioritize the policies against the spousal violence for increasing the survival rate of under five girls. Abused women may emotionally and physically get down and that will extensively discourage the child caring.

Boco (2014) in his research investigated the sex differentials in childhood mortality limited with a specific region of Africa, analyzing individual and multilevel level factors separately. Pattern of female survival rate shows the Longer-term persistence than male children due to poorer health conditions occurred as a result of existent issues related to biological factors.

Fertility Behavior of Woman

Family size with the increasing number of siblings tends to increase the risk of survival rates, specially associated with social and health related problems in later-born such as education attainment, adult health, job market outcomes. This provides the evidence of the higher risks in female birth order compared to male children in a family. Therefore, the parent's decision of fertility preferences is more influential on the quality of children. In despite of the later-born issues, first-born influences with low birth weights and nutritional problems (Barclay and kolk, 2013) and (Lundberg and Svaleryd, 2016)

Birth Spacing is also a more important character, when allocating resources for siblings. For the increased survival rate, it needs to devote much time as a mother. If there is a twin birth, mortality rate significantly increases. Higher the female literacy, it reduces the risk of mortality with increasing the birth interval by Makepeace and Pal (2006). Effect of birth spacing on infant and child mortality was discussed by Kembo and Ginneken (2009) in Zimbabwe. The study revealed that the risk of multiple birth is 2.08 higher than the singleton birth and the short birth interval results higher, the risk of mortality level. Then the society needs the good level of knowledge towards family planning.

Dissanayake (2000) examined, how inter-birth intervals affected by child mortality and breastfeeding. Lactation makes longer the postpartum amenorrhea period. 50s 60s onwards woman knew breastfeeding has a contraceptive effect against pregnancy, no affect found on birth intervals. In 1979 government introduced the family planning programs. Woman stated sexual intercourse before restating the menstruation. Nobles et al. (2015) tried to identify the relationship between mortality and Fertility after arising natural disaster. There was an increasing trend in fertility with affecting disaster among women who lost their children for the additional birth of one or two children.

2.3.2 Socio-Economic Factors

Mother's Education/ Literacy

Education leads for increasing the social value of woman as stated by Kaufmann and Cleland (1994). When she is strong, she possesses a healthy image with greater personality. It a good sign of moving with the world, understand the power of self-confidence and that will direct to hold a strong motherhood in the later life as a responsible character in the society. Ewbank (1994) extended his research into social, economic and behavioral factors with child survival. It was analyzed the way how social and cultural changes effects on health by observing individual and macro-level behavioral changes.

As a mother to interact with the available medical resources, literacy gained from their childhood schooling is effectively taken place in a higher level, research done by Vine et al. (1994). That is the result of reading performance given the ability of reading health messages irrespective of the quality of the school which was attended by the mother. In reducing mortality, female literacy programs effect in the way of long term as well as short term when achieving sustainable development goals. Pillai et al. (2013) studied the longitudinal association between child survival and the female literacy in two ways by focusing on women's empowerment and reproductive rights.

Shetty et al. (2014) examined the relationship between female literacy and child mortality. Results shows there is an inverse relationship between them. Educated woman build ups healthier children. It should be motivated to boost the female literacy in the society.

Woman's Rights in the Society

Women's employment empowers the women beside less freedom caused to higher level of mortality risks in certain kinds of works entitled. Basu et al. (1991) emphasized this by examining working mothers in South Asia. Working women experience higher mortality than non-working women despite the income level of the family. Women with wealthier households have the better environment for the easy living. At any moment, what is her requirement, it will be provided. There are no difficulties to access the best healthcare services and professionals within less time duration. That is a sign of reducing mortality risk as examined by Griffis (2012). For the divorced women, situation may change according to the resources she holds.

Low birth weight and prematurity may be the adverse effects of violence during pregnancy. Ahmed et al. (2006) studied the physical domestic violence against pregnant women on perinatal and childhood mortality in India, it was 18% among sample size of 2199 women. The risk of childhood mortality is 2.59 times higher compared to women who do not suffer violence. Rawlings and Siddique (2018) analyzed the same issues based on 32 developing countries as a whole by focusing on the legal framework against domestic violence of female victims and their victimized children.

2.3.3 Health/Environmental Factors

Medical Disorders of Mother/Child

As the prematurity and low birth weights are major causes of Neonatal mortality. Many studies were conducted to identify, how these issues are influenced on child mortality. It was proved that gastroenteritis, issues in respiratory system, Opium Intoxication, habits of smoking are severely affected on post-neonatal mortality. Then concluding analysis, it was recommended to special care in the infantile period for mother and the child by Sharifzadeh et al. (2008).

Preterm babies with prematurity is a global challenge to be achieved against the neonatal mortality and morbidity, a study results by Gamhewage et al. (2018). The worldwide mortality rate is approximately 50%-70% due to patent ductus arteriosus, surfactant deficient lung disease, hypotension and sepsis. Survivors will live only 45 days. This is an extreme situation.

Importance of the Food/Nutrition

Practice of infant breastfeeding influence the nutritional status of children that lows the risk of mortality when initializing breastfeeding and increasing the duration. Malnutrition depends on Mother's literacy and fertility behavior, a study by Al-Jawaldeh et al. (2018). Recent reports say that half of the deaths of under-five children are victims of malnutrition and mostly interconnected with diarrhea and acute respiratory infection then malaria as well, least association with measles. Rice et al. (2000) focused on how infectious diseases associated with the risk of child mortality due to malnutrition.

Considering, how the effects of soil type, as a measure of farm-level agricultural productivity on child morality, Hedefalk et al. (2017) concluded in their study with higher risk of mortality shows the areas of clay and sandy soils directly affected the nutrition of farmer's children. Laborer's children who lives in mixed soiled areas are away from the higher risk of dying.

Maternal malnutrition, low gestational weight are the key reasons for neonatal mortality and low birth weight of new born babies in developing countries, maintaining with existence of high fertility. Das (2015) emphasized that half of the world's population suffers from nutritional deficiencies under the common conditions of intrauterine growth restriction (IUGR), low birth weight, lack of protein-energy, chronic energy undernourishment, and micronutrient deficiencies.

Importance of the Medication for the Mother/Child

It is worst that children who are incompletely immunized are in danger with individual, community-level and state level factors as well. Adedokun et al. (2017). Government and non-government organizations are the responsible parties, to take the necessary actions when processing the policies to overcome these factors. Shann (2013) suggested that slight changes of routine immunization schedule may increase the child survival by thirty percent. Further, the BCG and measles vaccines reduce mortality, despite DTP may increase mortality, especially in girls. That is the main reason for child mortality.

When reducing child mortality and morbidity, it is considered Vitamin A and Zinc supplementations as a highly influential medication. Benn (2012) explored the combination of Vitamin A and Vaccines, concluded with low income countries are suffer from Vitamin A deficiency. Further current health policies should be revised, dealing with girls and boys separately. Koenig et al. (1990) proved reduction of mortality, as a result of measles vaccination becomes higher. Further it was found that 46% of reduction of death risk of vaccinated children compared to non-vaccinated children in Bangladesh and emphasized the importance of giving the priority for measles vaccination in WHO programs.

2.4 Social Responsibility of Reducing Child Mortality

In India, under-five children death has fallen down slowly with non-income factors over the past 10-20 years. Then there should be new strategical approaches for preventing childhood diseases, awareness of the health and nutritional cycle of mother & child, establishing state-specific disease patterns such as epidemic and demographic disorders (Claeson et al., 2000). In Bangladesh, Khan and Awan (2017) argued on eradication of risk of children death by suggesting to focus on unseen community living in the society. Moreover, child survival programs need to be planning, focusing on birth spacing. Urban rural disparity leads higher risk of problem.

Developing human welfare is playing a significant role in social growth of developing countries. In contrary, world-bank reports, reducing mortality, health spending doesn’t play a big role corresponding to income per capita of a country as stated by Hanmer et al. This study proves that pursing consistent improvements in health services in a strongly targeted robust health strategies, social policies will be effectively directed to achieve the saving of lives. As an example, immunization is a cost-effective intervention of targeted goals. Moreover, cost differences play an insignificant role in mortality regression on health spending.

2.5 Review of Classification Models

Statistical classification of the categorical outcome variables into observed groups is a widely used area in the modern world data science. Success of the achieving the study objectives has been depended on two statistical methods, discriminant analysis and logistic regression. Number of researchers had employed these two traditional statistical methodologies, in their studies in the field of child mortality.

2.5.1 Discriminant Analysis

Considering the determinants influenced on child well-being and survival, a study on child mortality in MENA countries by Ali (2005) had established a general discriminant model in identification of most discriminant variables over the few levels of analysis: household, individual and community level. The study observed that socio-economic and environmental factors have a significant effect on the child's household. The author, in addition, recommended that to compose an index for measuring the well-being of children with health, social, material, behavioral, emotional well-being etc.

Severe acute malnutrition is strongly linked with Diarrheal. Writing on the significance of the acute malnutrition on the child mortality, Attia et al. (2016) examined the hypothesis of the child mortality with severe acute malnutrition, which is associated and interconnected with diarrhea, enter pathogenesis, and systemic and intestinal inflammation. The authors noted that the observed determinants greatly influence on the death of children, aged 6-59 months in Malawi with severe acute malnutrition. The established discriminant model resulted that high systemic inflammation was directly related to the mortality and diarrhea, high calprotectin, and low SCFA production were indirectly related to the child mortality.

Thus, in another study Greenberg et al. (1963) stressing variation between perinatal death and survival of the neonatal period with the analysis of 30 determinants in North Carolina. Results of linear discriminant analysis model explained the 54% of the variation of the study objective and 39% of survival status of the infant. It was further established that bleeding complications, illness in present pregnancy and fetal distress constitutes a significant cause of perinatal deaths.

Life expectancy indicates the overall mortality level in a country as stated by world health organization. Several studies have examined the influence of life expectancy patterns on mortality. Higher mortality rates are accompanied with children, aged 0-5. Sufian (2013) supports this argument by analyzing the different socio-economic factors on life expectancy at birth based of national level data of selected countries in 2011. Standardized coefficients along with structure coefficients have been assessed with canonical discriminant analysis. The efficiency of determinants was measured by using Wilk's lambda, canonical correlation, plot of group centroids etc. The author found that the infant mortality is strongly correlated with the life expectancy at birth.

Garenne et al. (2003) contends that wealth index has a direct or indirect connection with the survival of children, aged 0-5. He proposed an analytical framework for the study of factors of child mortality in Morocco with analyzing DHS data, in 1992. Discriminant analysis employed 15 socio-economic variables for screening the families at higher risk of infant and child mortality. It was recommended that to reduce the significant risk of child mortality by introducing free health insurance system for the urban poor people.

2.5.2 Logistic Regression

In investigating the impact of various factors on child mortality, Islam et al. (2013) and Rahman et al. (2019) employed the logistic regression model to analyze the data from demographic and health surveys, in Bangladesh respectively, year 2007 and year

2014. Both studies supported the hypothesis that social, demographic health and environmental factors, negatively associate with the mortality. According to them, by enhancing the woman's education, expansion of birth intervals, proper breastfeeding practices are the solutions for controlling the issue of analysis. Community based health care centers must be attention more on maternal and child care activities in Bangladesh. Their study has also been supported by the previous study conducted by Hossain and Islam (2008) in Rajshahi district, Bangladesh with the analysis of 800 individual samples. He identified that the distribution of social benefits and the economic conditions in a community are significantly influenced on the child mortality patterns by modelling the socio-economic factors, identified.

In a study that investigated the impact of malaria on the prevalence of infant mortality in Ghana 2010-2015, Ahinakwa and oduro (2018) explained that the malaria does not contribute to the mortality and only the factor, duration of stay in hospital significantly effects on the mortality among the other independent variables.

In classification modelling, deep learning techniques like machine learning are superior to logistic regression modeling in identifying the significant determinants in under-five mortality. The evidence provided by Adegbosin et al. (2020) in his study with analyzing the data in low-income and middle-income countries, concluded that in classification of child survival, LR sensitivity=0.47, specificity=0.53 were lower than that of other modern analytical approaches.

Vidal-e-Silva et al. (2018) aimed to establish the factors associated with preventable child deaths using a systematic sample of 4,402 children from a population of 34,284 live births. Multiple hierarchical logistic regression models were used in analysis of four hierarchical models in socio-demographic factors. The findings showed that 62.1% of deaths of infants could have been prevented with better diagnostic and treatment procedures during the pregnancy of women and the preventable deaths were determinants of poverty and poor nutrition of both mother & child. Zewdie and Adjiwanou (2017) also carried out a study to highlight the impact of poverty and inequality on infant mortality in South Africa. The results from the multilevel logistic regression model revealed that demographic & socio-economic factors have significant effect on deaths of infants and at the individual level it is attributed from -24% to 15% change on the odds of death. Following inferences were made: as the first, education of woman and the standard living conditions are associated with lower rates of mortality, secondary, higher income inequality & HIV prevalence are associated with higher rates of mortality.

Aheto (2019), in his study attempted to develop a predictive model using single level binary logistic and the multilevel logistic regression models to identify the factors associated with under-five mortality in Ghana. Likelihood ratio test, VIF and ROC curves were generated to assess the predictive ability. The findings cited that family planning, maternal health education and regional gap in child health are strongly associated with the mortality.

Furthermore, the literature review found that socio-economic, health, demographic and environment covariates had consistent findings on under-five child mortality, employed with logistic regression model building in different geographical regions. Several studies proved that of significant influence on mortality are Bhatia et al. (2019), Adetoro and Amoo (2014), Geneti and Deressa (2014), Maniruzzaman et al. (2018), Fenta et al. (2020) and Tagoe et al. (2020). Addressing these issues will help to prevent child mortality in future by improving health care facilities of mother and child.

In developing countries, generating more precise and timely estimates in infant and child mortality requires more cost and substantially high methodological expertise. Garfield and Leu (2000) in their study, explained that the usefulness in using rapid estimation technique by establishing four preliminary models and then developed a logistic regression model with six social determinants for countries with complex humanitarian emergencies, which are not been able to produce the routine data.

2.6 Comparison of DA & LR Models

Childhood pneumonia is a major disease that cause in child mortality, aged between 2-59 months. Tuti et al. (2017) in his study focused on the determinants of non-severe pneumonia in the exploration of child mortality in Kenya, between the periods of 2014-2016. It is hypothesized that the effectiveness of clinical characteristics is associated with the risk of inpatient mortality for pneumonia. Machine learning techniques were used to analyze the top most risk factors and explored the decision curve analysis. The findings indicated that higher sensitivity in partial least squares discriminant analysis than that of logistic regression modelling in predicting mortality. Then it was concluded that inpatient care is strongly influenced on the study issue.

Several studies have compared the classification accuracy by comparing discriminant analysis and logistic regression in modeling infant and child mortality. As an example, Hanley (1983) has studied the uses of multivariate analysis with the application of mortality data, children aged 0-5. Maniruzzaman et al. (2018) in their study, established a model with neonatal and child mortality with the application of using chi-square ($\chi 2$) statistic and multiple logistic regression. Finally, it was recommended to restructure the model with applying discriminant analysis in the future.

Child birth is considered high-risky situation of both mother and child when potential complications are arisen. Balogun et al. (2015) proposed his study to establish a statistical model using discriminant analysis and logistic regression in predicting the risk of the mode of delivery that is associated with the mortality of mother and the child for analyzing the 184 samples. It has considered to types of mode of delivery: natural births and the caesarian child births. Results revealed that correctly classified 64.6% of natural births and 64.7% of caesarian births in discriminant analysis and that 76.8% of natural births and 52.9% of caesarian births in logistic regression modelling. Then it was concluded that with better classification accuracy in both techniques and the similar statistically significant coefficients in modelling the child survival.

Worldwide, many neonatal deaths are occurred due to a major disease: sepsis. Thakur and Pahuja (2019) decided to establish a framework for the performance comparison of multiple discriminant analysis and logistic regression with modeling the factors influenced on neonatal sepsis. The established model identified that inaccessibility of hospitals and absence of laboratories are the reasons for most of the neonatal deaths in less developed countries.

Jayadevan (2016) conducted his study to identify the factors influenced on deaths of under-five children among the countries of Asia within the period of 1995-2013 with comparative statistical analysis between discriminant analysis and logistic regression modelling. Finally, it was concluded that significant increase in discriminating variables, reduce the child mortality.

2.7 Summary of Chapter 2

Many studies have been conducted in assessing the relationship between deaths of under-five children with identified factors in terms of demographic level, social-economic level and environment-health level with the country level variations. However, there are various gaps as explained above. However, the extensive review is useful to plan this study. The necessity of comparative analysis incorporating various statistical methods in modelling child mortality was recognized.

CHAPTER 3

MATERIALS & METHODOLOGY

This chapter provides the detailed description of the secondary data used along with statistical methodologies applied and describes the statistical methods that will be expecting to incorporate when measuring the relationship between dependent and independent variables further.

3.1 Population & Sample

Women's questionnaire of SLDHS, 2016 was selected as the main source of gathering information for this study. There were representatives of 28,720 housing units and 18,302 female participants and ever-married women of reproductive age 15-49 who were successfully interviewed by the officers appointed by the census & statistical department in Sri Lanka by covering the entire country within the period of May, 2016 to November, 2016. The questionnaire covered the detailed information about their children too, born after January 2011 who were under five years.

3.2 Sampling Procedure

Multistage stratified sampling design was used as the sampling framework with 2 stage of sampling designs. The questionnaire contained two main parts, 'households' and 'women & children'. There were over 1500 variables in the dataset. Each one had the potential values in numerical & alphabetical formats accordingly for easy of collecting information in the questionnaire. This study contains only 10 variables for formulating the relationship between dependent and independent variables.

3.3 Methods of Data Collection and Analysis

For the purpose of modelling the under-five child mortality, secondary data were obtained from the National Demographic Health Survey, 2016 conducted by Census

& Statistical Department in Sri Lanka. The dataset has been divided into five subsets of separate categories, children, households, kids, persons and women for protecting the privacy of individual samples and supporting the efficient path for the data analysis in the future.

Quantitative and qualitative (dummy-coded) data were analyzed by using two classification techniques: discriminant analysis and binary logistic regression with standard statistical package: SPSS. Information related to the women & Children were chosen to conduct the analysis aligned with the study objectives.

Finally, results were obtained by executing same set of selected determinants using both statistical methodologies. Study results were interpreted according to the specifications of selected analytical methodologies based on the assumptions. Efficiency analysis between two methodologies was performed by obtaining the classification accuracy, sensitivity & specificity and ROC curves.

3.4 Conceptual Framework

This conceptual framework incorporates three major process levels in terms of risk factors, direct effects and outcome variables. According to the model, direct effects are the impacts of individual's health status and that makes the linkage between the risk factors and the individual's status of survival or death. Direct effect comprises the diseases/infections, which were leading to the mortality of children.

Many studies are available in focusing on studying the frameworks for identifying the determinants on child mortality. Few of them are Mosley and Chen (1984), Hill (2003) and Masuy-Stroobant (2001). With the studies incorporated, they have established hypothesis that how different factors are influenced on deaths or survival of children by proposing an analytical framework for identification of most influential factors.

Risk factors are the direct consequences of factors, which are supposed to be measured in relation to the mortality status of the children. In this study, three set of factor level entries incorporated in terms of health-environmental, socio-economic and demographic levels, thus the individual components were selected accordingly. Figure 3.1 illustrates the conceptual framework of the system.

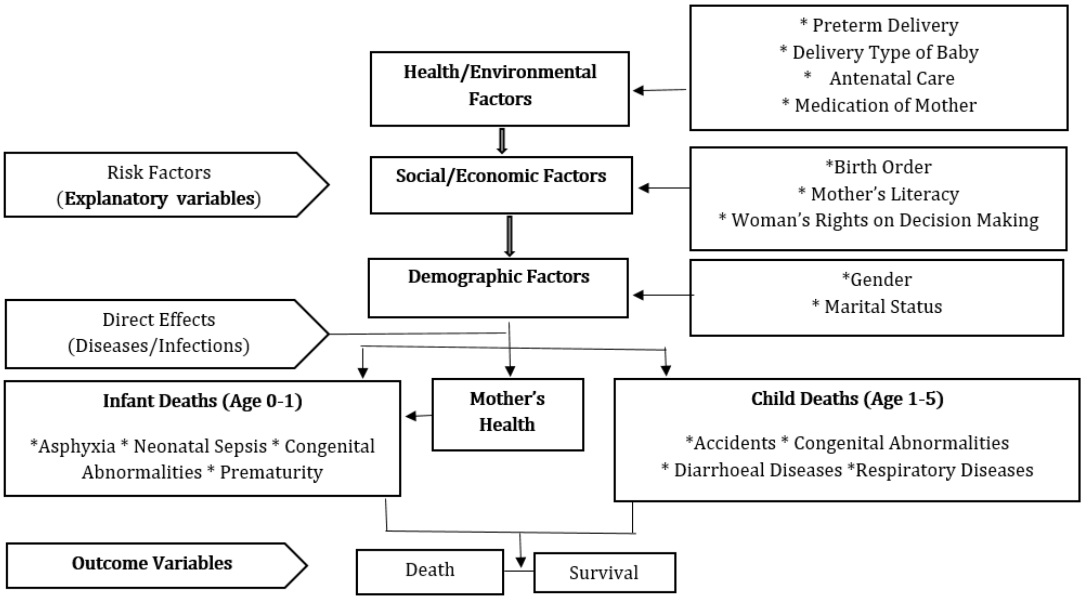


Figure 3.1: Conceptual Framework of the Research

3.5 Selection of Variables

The statistical process of variable selection was based on the assumptions of both binary logistic regression and discriminant analysis models. Binary logistic regression allows any kind of independent variables such as continuous, discrete, categorical and dichotomous categorical variable as the dependent variable. In the discriminant analysis, for the assumption of multivariate normal distribution, it generally allows only the continuous or ordinal variables as the independent variables and categorical & dichotomous are accepted as the dependent variable (Antonogeorgos et al., 2009).

When there are continuous variables for independent variables, discriminant analysis works well. Discriminant correspondence analysis is the most appropriate technique for the categorical independent variables. Discriminant correspondence analysis is an extended version of discriminant analysis. Performing discriminant analysis after converting categorical variables into dummy variables is an alternative method of analysis. As stated by Krzanowski(1986), many users claimed that SPSS automatically convert the categorical variables into dummy variables, hence linear discriminant analysis can be performed and it is a form of canonical correlation analysis, when there are mixed case variables.

In this study, all the dichotomous variables were assigned ‘0’ and ‘1’. Such details are shown in Table 3.1.

Table 3.1: Detailed Description of the selected variables

Variables	**Type**	**Definition**	**Description**
Independent Variables:			
Gender (X1)	Categorical	Gender of Child	0 - Female 1 – Male
Birth Order (X2)	Continuous	Order of the Child Birth in a family	
Marital Status (X3)	Categorical	Whether person is married	1 - Married 2 - Living Together 3 - Widow/ Divorced/ Separated
Mother's Literacy (X4)	Categorical	Mother’s ability of read and write	1 - Cannot Read at all 2 - Able to Read only Part of the Sentence 3 - Able to Read whole Sentence 4 - No Card with Required Language 5 - Blind / Visually Impaired
Status of Antenatal Cate (X5)	Categorical	Whether routine health control is obtained	0 – No 1 – Yes
Status of Vitamin/Medicine Usage (X6)	Categorical	How often taking vitamins/ medicines	1 – Daily 2 - Less Often 3 – Other
Delivery Type (X7)	Categorical	Whether the type of medical assistance during delivery of baby is Normal	0 – No 1 – Yes
Pregnancy Duration (X8)	Continuous	Number of months of the pregnancy	
Decision Making Ability (X9)	Categorical	Who is taking the decisions on earnings	1 – Woman 2 - Husband / Partner 3 - Husband / Wife jointly 4 - Husband / Wife has no earnings 5 – Other

Table 3.1: Detailed Description of the selected variables (Continued)

Variables	Type	Definition	Description
Dependent Variable:			
Status of Child Mortality (Y)	Categorical	Whether child/ Infant is dead	0 – No 1 - Yes

Dependent Variable

Status of child mortality was selected as the dependent variable which is dichotomous and the Outcome consist of the probability of dying children aged between 0 - 5. Thus, the status of survival represents as ‘0’ and the death represents as ‘1’ by assessing the risk of deaths during first few years of a child in Sri Lanka.

Independent Variable

Independent variables are established in the individual level, community level and household level by reviewing the previous studies related to the same study problem. Table 3.1 demonstrated, nine variables, which were selected as the predictor variables in this study. Of the nine variables two variables, X2 and X8 are continuous variables while the balance seven variables are categorical variables.

3.6 Key Measurements

Infant Mortality Rate (IMR)

Infant mortality rate constitutes two parts: neonatal mortality rate and post-neonatal mortality rate. However, Infant mortality rate refers the probability of death of children from the birth to first year birthday (0-11 months) occurred per thousand live births. The formula can be defined as follows:

$$\mathrm{IMR} = \left(\frac{\text{Number of Child Deaths,aged } 0\!-\!1 \text{ year for a Specified Period}}{\text{Total Number of Live Births for the Same Period}}\right) * 1{,}000 \qquad (3.1)$$

Child Mortality Rate (CMR)

The child mortality rate is calculated by the number of deaths of children aged between the first birthday and the fifth birth day (12-59 months) at a given period of time, to the total number of live births multiplied by 1,000. The formula can be defined as follows:

$$\text{CMR} = \left(\frac{\text{Number of Child Deaths,aged 1−5 year for a Specified Period}}{\text{Total Number of Live Births for the Same Period}}\right) * 1{,}000 \qquad (3.2)$$

3.7 Dummy Coding of Selected Variables

Dummy variables and binary variables are some interchangeably used terms in statistical analysis. Dummy coding is a method of quantifying categorical variables in regression analysis. In this analysis, it was used two continuous variables and seven dummy coded categorical variables.

When performing binary logistic regression, SPSS software treats the categorical variables, hence no need of using the dummy coding. In discriminant analysis too, continuous and multilevel ordered categorical variables are allowed in the discriminating process.

3.8 Univariate Test Statistics

3.8.1 Chi-Square Test

As a non-parametric test, Chi-square test is performed to test the significance of interaction between the observed and expected frequencies. In this analysis, it was performed for obtaining the significant association between two categorical independent variables. (Maiprasert and Kitbumrungrat, 2012) It is implemented to test following hypothesis:

H_0: The sample has been drawn from population following a specified distribution.
H_1: The sample has not been drawn from population following a specified distribution.

Pearson Chi-square goodness of fit test can be expressed as:

$$x^2 = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i} \qquad (3.3)$$

Where,
O_i = observed values

E_i = expected values

X^2 = chi-square value

According to the hypothesis tested, null hypothesis will be rejected, if the chi-square value exceeds critical value and it accepts the null hypothesis, if the chi-square value is less than that.

3.8.2 Two Sample t-test

As a parametric test, two sample t-test is used to test the observed difference between two population means under the null hypothesis shown below.

Thus, null hypothesis and alternative hypothesis are:

$H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$

$H_0 : \mu_1 - \mu_2 = 0$ against $H_1 : \mu_1 - \mu_2 \neq 0$

The t-test statistics is given by the formula, (3.4).

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_1}} \tag{3.4}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where,

$\bar{x}_1$= sample mean of the first population

$\bar{x}_2$= sample mean of the second population

μ_1= true mean of the first population

μ_2= true mean of the second population

$\sigma_{\bar{x}_1 - \bar{x}_1}$= standard error of the difference between the two sample means

Under,

H_0, $t \sim T_{n1+n2-2}$

3.9 Discriminant Analysis

Discriminant analysis is applied to develop a linear function to discriminate the k groups using a linear function of the selected variables and to classify a new observation to one of those groups.

The form of linear discriminant function is:

$$D = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \ldots + \beta_k X_{ik} \tag{3.5}$$

Where,

D = discriminate score of discriminant function of the i^{th} observation β_i = discriminant coefficient or weight of the i^{th} variable, X_i = i^{th} observed variable, β_0 = constant

3.9.1 Wilk's Lambda Statistic

$$SS_T = \sum_{L=1}^{K} \sum_{i=1}^{N_L} (X_{Li} - M)(X_{Li} - M)' \tag{3.6}$$

$$SS_B = \sum_{L=1}^{K} \sum_{i=1}^{N_L} (X_{Li} - M_L)(X_{Li} - M_L)' \tag{3.7}$$

Wilks' lambda as an inverse quality yardstick evaluates the level of significance of the function of the discrimination. As the first step, cross products matrices are calculated for between group differences (eq.3.7) [SS_B-between sums of squares and products] and within group differences (eq.3.8) [SS_W-within sums of squares and products] (Mahmoud, 2008).

Assume, k = number of groups, N_k = observations of the k^{th} per group, N = total number of observations,

$$SS_T = SS_B + SS_W \tag{3.8}$$

Then eigenvector, V can be defined as

$$V = s_W^{-1} s_B \tag{3.9}$$

Where, SS_T = total observed variability of the system

$$\Lambda = \frac{|S_W|}{|S_B + S_W|} \tag{3.10}$$

Testing the significance of the independent variable, Wilks' lambda (eq. 3.11) is used in the ANOVA F test of analysis. Lambda ranges from 0 to 1 that 1 means the all group means are similar and lesser contribution to the discriminant function. (Poulsen and French) Wilks' lambda is defined as follows:

$$\Lambda = \frac{|S_w|}{|S_T|} \tag{3.11}$$

$$= \det |(S_W+S_B)^{-1} * S_E| = \prod_{j=1}^{m} \frac{1}{1+\lambda j}$$

Where, λ_j = the j^{th} eigenvalue corresponding to the eigenvector and m = the minimum of k-1 and p.

3.9.2 Likelihood Function

Likelihood function (eq. 3.12) is a goodness of fit test in statistical modeling. It is estimated the likelihood function for calculating the conditional probability. Simply it can be denoted using joint probability density function f(x|θ) :

$$L(\theta|x) = f(x|\theta) \tag{3.12}$$

Maximum likelihood is a method of finding the value of one or more variables for a given dataset. By maximizing the expression in order to find the probability distribution, an estimate of likelihood function (eq. 3.13) can be represented as $\pi(G_i|D)$ (Abdulqader, 2015):

$$\pi(G_i|D) = \frac{\pi(D|G_i)\pi(G_i)}{\sum_{j=1}^{n} \pi(D|G_j)\pi(G_j)} \tag{3.13}$$

For the predictor i, objects are belonging to different groups based on the Z score value.

3.9.3 F-Test Statistic

To test the statistical significance of regression sums of square, overall F-statistic is commonly applied by the statisticians. In the discriminant function analysis, wilks' lambda is assessed by F test for the significance of all the independent parameters. When n is equal,

$$y = \Lambda^{\frac{1}{s}} \tag{3.14}$$

Then,

df_{effect} = number of groups minus one, (k-1)

$$s = \sqrt{\frac{P^2(df_{effect})^2 - 4}{P^2 + (df_{effect})^2 - 5}} \tag{3.15}$$

df_{error} = number of groups times (n-1): k(n-1)

df_1 = p(df_{effect})

$$df_2 = s\left[(df_{error}) - \frac{P - df_{effect} + 1}{2}\right] - \left[\frac{p(df_{effect}) - 2}{2}\right] \tag{3.16}$$

For unequal of n, df_{error} = N-k

Then F-statistic (eq.3.17) is obtained in MANOVA. (Poulsen and French)

$$F_{approximate}(df_1, df_2) = \left(\frac{1-y}{y}\right)\left(\frac{df_2}{df_1}\right) \tag{3.17}$$

3.9.4 Box-M Statistic

Box-M Statistic (eq. 3.18) is a test for homogeneity/equality of covariance matrices of the explanatory variables of the dataset. That means the generalized variances of each group of datasets. The test is responsive limitedly, hence it is exceptionally verified the normality assumption according to the fundamentals of discriminant analysis rather than the inequality of covariance matrices. (Bhaumik) Then the hypothesis is derived to get the non-significant of test results with equal covariance matrices of each group,

In 1949, Box suggested a likelihood ratio for the verification of hypothesis of equality of covariance matrices as given below (Friendly and Sigal, 2018).

$$M = (N - g)\ln|S_P| - \sum_{i=1}^{g}(n_i - 1)ln\,|S_i| \tag{3.18}$$

Where, N = total sample size, S_p= pooled covariance matrix

3.9.5 Mahalanobis Distance

Mahalanobis distance (eq. 3.19) is used the group means and variances for each variable. It is useful for measuring learning curves, serial position effects and group profiles. (Mahmoud, 2008) The formula is defined with i and j, between two correlate points as:

x_i and y_i are taken as the two correlated variables,

$$MSD_{ij}^2 = \frac{1}{1-r^2}\left[\frac{(x_{i1}-y_{j1})^2}{s_1^2} + \frac{(x_{i2}-y_{j2})^2}{s_2^2} - \frac{2r*(x_{i1}-y_{j1})}{s_1}\frac{(x_{i2}-y_{j2})}{s_2}\right] \qquad (3.19)$$

Where, s_1^2 and s_2^2 sample variances, r = correlation coefficient

When it is extended further,

$$MSD_{ij}^2 = (x_i - y_i)' S_i^{-1}(x_i - y_i) \qquad (3.20)$$

Where, s^{-1} = inverse covariance matrix.

3.10 Binary Logistic Regression

Logistic regression analysis applies for binary or multinomial (ordinary type) response data to be investigated the relationship between the explanatory variables and the log odd ratio of an event.

The logistic regression model can be expressed as follows:

$$Log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_{2+\dots} + \beta_k x_K \qquad (3.21)$$

$$= \beta_o + \sum_{i=1}^{k} \beta_i X_i$$

Where,

p = probability of success, (1-p) = probability of failure, β_k = value of coefficient of factor x, x_k = value of the independent variable, n = 1,2,3,4……, k, β_o = intercept

Logit function of the probability is the logarithm of the odds. When the observation of the factor x having two levels:

$$Logit(p(x)) = Log\left(\frac{p(x)}{1-p(x)}\right) = \beta_0 + \beta_1 x \tag{3.22}$$

In the binary logistic regression, when logit model is used to model the dichotomous variables,

$$p(x) = \frac{e^{(\beta_O+\beta_1 x)}}{1+e^{(\beta_O+\beta_1 x)}} \tag{3.23}$$

3.10.1 Odds Ratio

Odds ratio is a measure of association between event and outcome, which represents the odds of an event happing compared to the absence of that exposure with the particular outcome.

The odds of an event occurring are as follows:

$$odds = \frac{p}{1-P} \tag{3.24}$$

Probability of occurrence can be estimated by using below logistic regression model in the binomial distribution.

$$P(Y_i = 1|X_i) = \frac{odds}{1+odds} = \frac{e^{(\beta_O + \sum_{i=1}^{k} \beta_i X_i)}}{1+ e^{(\beta_O + \sum_{i=1}^{k} \beta_i X_i)}} = \frac{1}{1+e^{(-\beta_O - \sum_{i=1}^{k} \beta_i X_i)}} \tag{3.25}$$

3.10.2 The Wald Test

Wald test is the function of difference between the maximum likelihood estimate and the hypothesis value (Ahmed, 2017). Wald test statistic is as a function:

$$W^2 = \frac{(\hat{B}-\beta_O)^2}{\widehat{Var}(\hat{B})} \sim x_1^2 \tag{3.26}$$

Where,

x_1^2 = Distribution with one degrees of freedom

This test is used to determine the certain predictor variable, significant or not by testing hypothesis, then rejects null hypothesis if corresponding coefficient being zero.

3.10.3 The Hosmer-Lemeshow Goodness of Fit Test

The Hosmer - Lemeshow test is a goodness of fit test for logistic regression. It can be given by the statistics (Ahmed, 2017),

$$X^2 = \sum_{i=1}^{n} \frac{(o_i - n_i p_i)^2}{n_i p_i (1 - p_i)} \tag{3.27}$$

Where,

n_i = total number of cases in the i^{th} group

o_i = number of event outcomes in the i^{th} group

p_i = average estimated probability of an event outcome for the i^{th} group

Output will be resulted with chi-square value with p-value. When p-value is less than 5%, it means model is poorly fit with data.

3.10.4 Pearson Chi-Square

Chi-square test is performed to test the significance of interaction between the observed and expected frequencies. (Maiprasert and Kitbumrungrat, 2012) It is implemented to test following hypothesis:

H_0: The sample has been drawn from population following a specified distribution.

H_1: The sample has not been drawn from population following a specified distribution.

Pearson Chi-square goodness of fit test can be expressed as:

$$x^2 = \sum_{i=1}^{r} \frac{(O_i - E_i)^2}{E_i} \tag{3.28}$$

Where,

O_i = observed values

E_i = expected values

X^2 = chi-square value

According to the hypothesis tested, null hypothesis will be rejected, if the chi-square value exceeds critical value and it accepts the null hypothesis, if the chi-square value is less than that.

3.10.5 Maximum Likelihood Estimate

Probability distribution of the dependent variable is used to form the maximum likelihood estimate. MLE is a template to fit classification model more generally (Maiprasert and Kitbumrungrat, 2012). It is expressed by this equation:

$$G^2 = -2[\ln L_p - \ln L_0] \tag{3.29}$$

Where,

df = p

L_p = likelihood of constant value and group of independent p-value

Lo = likelihood of only constant value

$$Pr\left(Y_0 = 1|\hat{B}\right) = \hat{\pi}_0 = \frac{1}{1+e^{-x_0\hat{B}}} \tag{3.30}$$

Logarithm of log likelihood function:

$$ln\, L = \sum_{i=1}^{n} y_i\, ln\, P_i + \sum_{i=1}^{n}(1 - y_i)\, ln(1 - P_i) \tag{3.31}$$

3.11 Model Evaluation

3.11.1 Analysis of Frequency Table

Contingency table is a two-way frequency table in a matrix format, which is used to examine the relationships between categorical variables as a collection of cell counts. Contingency table of joint and marginal probabilities is drafted as following Table 3.2.

Table 3.2: Contingency Table

		N					
		1	2	.	.	n	
M	1	a_{11}	a_{12}	.	.	a_{1n}	$a_{1.}$
	2	a_{21}	a_{22}	.	.	a_{2n}	$a_{2.}$
	.	.	.	.	.	.	.
	.	.	.	.	.	.	.
	M	a_{m1}	a_{m2}	.		a_{mn}	$a_{m.}$
		$a_{.1}$	$a_{.2}$	.	.	$a_{.n}$	

3.11.2 Estimate of Error Rate

Apparent Error Rate (AER) and Apparent Correct Classification Rate (ACCR) can be defined as the two criteria for evaluating the performance of statistical model building. Two terms are derived from equations as follows (Elgohari, 2017):

$$\mathrm{AER} = \frac{n_{12}+n_{21}}{n_1+n_2} \tag{3.32}$$

$$\mathrm{AER} = \frac{n_{12}+n_{21}}{n_{11}+n_{12}+n_{21}+n_{22}} \tag{3.33}$$

$$\mathrm{ACCR} = \frac{n_{11}+n_{22}}{n_1+n_2} \tag{3.34}$$

Where, n represents the number of observations, $n_1 = n_{11} + n_{12}$ and $n_2 = n_{21} + n_{22}$

3.11.3 Error Classification

Hypothesis testing is performed in the error classification, according to the criteria of Type I error and Type II error.

Table 3.3: Representation of Error Classification

		Predicted Value	
		TRUE	**FALSE**
Actual Value	**TRUE**	True Positive (Correct Inferences)	Type II Error (False Negative)
	FALSE	Type I Error (False Positive)	True Negative

3.11.4 Sensitivity/ Specificity

For the evaluation of two methods, sensitivity and specificity were used as a measurement of performance by following formulas (EL-habil & El-Jazzar, 2014).

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{3.35}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{3.36}$$

Where, TP: True Positive. TN: True Negative, FP: False Positive and FN: False Negative with prediction classes 1 or 2.

Table 3.4 illustrates the sensitivity/specificity in the tabular form.

Table 3.4: Representation of Sensitivity/Specificity

		Predicted Value		
		TRUE	**FALSE**	
Actual Value	**TRUE**	True Positive (TP)	False Negative (FN)	**Sensitivity**
	FALSE	False Positive (FP)	True Negative (TP)	**Specificity**
		Positive Predictive Value (PPV)	**Negative Predictive Value** (NPV)	

Accuracy can be estimated using sensitivity and specificity as follows:

$$\text{Accuracy} = (\text{Sensitivity}) * (\text{Prevalence}) + (\text{Specificity}) * (1\text{- Prevalence}) \tag{3.37}$$

Where, prevalence indicates the probability of status of child mortality at each point of considered.

3.11.5 Receiver Operating Characteristics (ROC) Curve

ROC curve is the graphical representation of the sensitivity and specificity, which is plotted sensitivity against the (100% - specificity) at different cut-off points. For the analysis of classification accuracy, several curves are plotted on the same axis (Montgomery et al., 1987).

When comparing few statistical models parallel, the larger the area enclosed to a particular model; it is defined as the model with better classification ability. Area Under Curve (AUC) is calculated at the 95% confidence interval. AUC is ranged between 0.5 to 1.

CHAPTER 4

ANALYSIS OF RISK FACTORS

This chapter provides the basic statistical analysis of the selected risk factor variables and their association with the status of mortality of children using univariate test statistics for independent variables. The response variable is binary independent variable.

4.1 Behaviour of Risk Factors

Risk factors were formed based on the demographic features, personal attributes and the natural behaviours of the focused group of characters in understanding the group difference of the independent variable. Set of variables were tested for the normality first and homogeneity of covariance at model building stage. Table 3.1 represents the detailed description of the predictor variables along with the dependent variable.

The classification of analysis is predicted whether the mortality status of children would be resulted positive or negative. Therefore, the status of child mortality has been classified as either survival or death. Assigned 'No' as the survival of children and 'Yes' as the death of children, aged between 0-5.

4.1.1 Distribution of the Selected Variables

Table 4.1 and table 4.2 present the distribution of all the selected variables of the research analysis, including the detailed description of the frequency and the percentage distribution. The survey had a total population of 10,589 sample size of individuals of children in the category of aged 0-5.

Table 4.1: Frequency Distribution of Response Variable

Response Variable	Description	Frequency	%
Status of Child Mortality (Y)	No	9806	92.6
	Yes	783	7.4

More than 90% of population belonged to the category of living children, it was exactly 9806 number of individuals. During the study period, the child deaths were constituted 783 individuals, the corresponding percentage of 7.4% of the total dependent population.

Table 4.2: Frequency Distribution of the Categorical Variables

Independent Variables	Description	Frequency	%
Gender of Child (X1)	Female	5097	48.1
	Male	5492	51.9
Marital Status (X3)	Married	9963	94.1
	Living Together	340	3.2
	Widow/ Divorced/ Separated	286	2.7
Mother's Literacy (X4)	Cannot Read at all	591	5.6
	Able to Read only Part of the Sentence	668	6.3
	Able to Read whole Sentence	9286	87.7
	No Card with Required Language	18	0.2
	Blind / Visually Impaired	26	0.2
Status of Antenatal Cate (X5)	No	2905	27.4
	Yes	7684	72.6
Status of Vitamin/Medicine Usage (X6)	Daily	4927	46.5
	Less Often	168	1.6
	Other	5494	51.9
Delivery Type (X7)	No	4261	40.2
	Yes	6328	59.8
Decision Making Ability (X9)	Woman	1202	11.4
	Husband / Partner	1516	14.3
	Husband / Wife jointly	6123	57.8
	Husband / Wife has no earnings	74	0.7
	Other	1674	15.8

4.2 Risk Factors influenced on Childhood Mortality

4.2.1 The Relationship between Mortality and Gender

The following hypothesis was tested using chi-square analysis and the results are shown in Table 4.3

H_0: No Significant association between mortality status and the gender.

H_1: Significant association between mortality status and the gender.

Table 4.3: Association between the status of mortality and gender

		Mortality Status				
		No		Yes		
		Count	**%**	**Count**	**%**	**Total**
Gender of Child (X1)	Female	4783	93.8	314	6.2	5097
	Male	5023	91.5	469	8.5	5492
Total		9806	92.6	783	7.4	10589

Pearson Chi-Square Test Statistics = 21.853 (p = 0.000)

Based on the results in Table 4.3, it can be concluded that there is a significant association between gender and child mortality as the observed chi-square test statistics is significant (p = 0.00). In other word, gender is significantly influenced on child mortality. Thus, it can be concluded that mortality rate of male (8.5%) is significantly higher than mortality rate of female (6.2%).

4.2.2 The Relationship between Mortality and Birth Order

Table 4.4: Association between the status of mortality and birth order

Mortality Status		**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
Birth Order (X2)	No	9806	1.94	0.993	0.010
	Yes	783	1.95	1.196	0.043

t-test Statistics = -0.273 (p-value = 0.785)

Results of Table 4.4 indicates that the comparison of birth order with child survival confirmed a significantly lower mean value than that of child deaths. Finally, it is illustrated that there is no significant association between birth order and status of child mortality as the p-value is greater than 0.05 under the assumption of equal mean of two groups of response variable.

4.2.3 The Relationship between Mortality and Marital Status

Table 4.5: Association between the status of child mortality and marital status

		Mortality Status				
		No		Yes		
		Count	**%**	**Count**	**%**	**Total**
Marital Status (X3)	Married	9283	93.2	680	6.8	9963
	Living Together	322	94.7	18	5.3	340
	Widow/ Divorced/ Seperated	201	70.3	85	29.7	286
Total		9806	92.6	783	7.4	10589
Pearson Chi-Square Test Statistics = 215.083 (p = 0.000)						

Table 4.5 illustrates the significant association between the status of child mortality and the marital status of mother as the observed chi-square statistic (215.0.083) is significant at 95% confidence interval. Mortality rate for the group of married women, 6.8% is not significantly different from the mortality rate of living together group. 5.3%. However, it is clear that the mortality rate for widow/divorced/separated group (29.7%) is very high and it is significantly higher than the corresponding rate of other two groups.

4.2.4 The Relationship between Mortality and Mother's Literacy

Table 4.6 reported the percentage difference been the child deaths and child survivals in Sri Lanka, which were belonged to the five levels of differences among mother's literacy. As chi-square test statistic (404.351) is significant at the 5% level it can be concluded that there is significant association between the status of child mortality and mother's literacy.

Table 4.6: Association between the status of mortality and mother's literacy

		Mortality Status				
		No		Yes		
		Count	**%**	**Count**	**%**	**Total**
Mother's Literacy (X4)	Cannot Read at all	454	76.8	137	23.2	591
	Able to read only Part of the Sentence	549	82.2	119	17.8	668
	Able to read Whole Sentence	8769	94.4	517	5.6	9286
	Can read foreign Languages	18	100.0	0	0.0	18
	Blind/Visually Impaired	16	61.5	10	38.5	26
Total		9806	92.6	783	7.4	10589

Pearson Chi-Square Test Statistic = 404.351(p = 0.0000)

The percentage of mortality for blind/visually impaired woman (38.5%) is much higher than that of other three groups. It should be noted that there is no mortality at all for woman who can read foreign language, though the number is small.

4.2.5 The Relationship between Mortality and Status of Antenatal Care

Table 4.7: Association between the status of mortality and status of antenatal care

		Mortality Status				
		No		Yes		
		Count	**%**	**Count**	**%**	**Total**
Status of Antenatal Care (X5)	No	2172	74.8	733	25.2	2905
	Yes	7634	99.3	50	0.7	7684
Total		9806	92.6	783	7.4	10589

Pearson Chi-Square Test Statistics = 1860.189 (p = 0.000)

Results of the Chi-Square test in table 4.7 confirmed that at the status of the mortality is significantly associated with the status of antenatal care as the p-value is less than 5%. It is indicated that the status of antenatal care significantly influences on child survival or child deaths. There is a significantly higher rate of mortality, 25.2% for woman who don't know about the antenatal care process.

4.2.6 The Relationship between Mortality and Status Vitamin/Medicines Usage

As explained above, the results in table 4.8 clearly indicates that the status of child mortality is significantly influenced by the status of vitamin/medicine usage. The child mortality is almost zero among the women who use vitamin or medicine either daily or less often. The corresponding rate for those who do not use vitamin or medicine at all (13.7%) is significantly much higher than that of other two rates.

Table 4.8: Association between the status of mortality and status of vitamin/medicine usage

		Mortality Status				
		No		Yes		
		Count	**%**	**Count**	**%**	**Total**
Status of Vitamin/Medicine Usage (X6)	Daily	4894	99.3	33	0.7	4927
	Less Often	168	100.0	0	0.0	168
	No	4744	86.3	750	13.7	5494
Total		9806	92.6	783	7.4	10589

Pearson Chi-Square Test Statistic = 652.874(p = 0.0000)

4.2.7 The Relationship between Mortality and Delivery Type

Since chi-square test statistics (968.393) is significant at 5% significance level, it is concluded that there is a significant association between the status of mortality and delivery type of the babies. According to the results exhibited (Table 4.9), it can be seen that 17.0% of child deaths has occurred, when the type of delivery is not normal. It is obvious that the above percentage is significantly higher than the corresponding rate in normal delivery.

Table 4.9: Association between the status of mortality and delivery type

		Mortality Status				
		No		Yes		
		Count	**%**	**Count**	**%**	**Total**
Normal Delivery (X7)	No	3535	83.0	726	17.0	4261
	Yes	6271	99.1	57	0.9	6328
Total		9806	92.6	783	7.4	10589

Pearson Chi-Square Test Statistics = 968.393 (p = 0.000)

4.2.8 The Relationship between Mortality and Pregnancy Duration

As explained by the table 4.10, comparison of child survival with pregnancy duration, illustrates the significantly higher mean value than that of the comparison between child deaths with pregnancy duration. Finally, it was confirmed that there is a significant relationship between status of child mortality and pregnancy duration under the assumption of equal mean of two groups of response variable.

Table 4.10: Association between the status of mortality and pregnancy duration

Mortality Status		N	Mean	Std. Deviation	Std. Error Mean
Pregnancy Duration (X8)	No	9806	8.99	0.435	0.004
	Yes	783	8.66	0.967	0.035

t-test Statistics = 17.814 (p-value = 0.000)

4.2.9 The Relationship between Mortality and Decision-Making Ability

Table 4.11: Association between the status of mortality and decision-making ability

		Mortality Status				
		No		Yes		
		Count	%	Count	%	Total
Decision Making Ability	Woman	1083	90.1	119	9.9	1202
	Husband	1412	93.1	104	6.9	1516
	Husband/ Wife Jointly	5704	93.2	419	6.8	6123
	Husband/ Wife has no earnings	67	90.5	7	9.5	74
	Not Applicable	11	91.7	1	8.3	12
	Other	1529	92.0	133	8.0	1662
Total		9806	92.6	783	7.4	10589

Pearson Chi-Square Test Statistic = 15.745(p = 0.008)

The chi square test statistic in table 4.11 (15.745) is significant (p = 0.008) confirming that there is a significant association between the status of child mortality and the decision-making ability. The lowest percentage was observed when the decision is taken by either husband alone or jointly by women and husband. In contrast the highest percentage was observed when the decision is taken by women alone.

4.3 Summary of Chapter 4

The variables such as, gender, marital status, mother's literacy, status of antenatal care, status of vitamin/medicine usage, delivery type, pregnant duration and decision-making ability are significantly influential variables on the status of mortality of children between 0-5 years of age. Birth order was omitted as the p-value > 0.05 at the 95% confidence level. Thus, those significant variables can be easily used to develop the binary logistic function and a discriminant function to separate the two groups in the next chapter.

CHAPTER 5

DEVELOPMENT OF LINEAR DISCRIMINANT MODEL

This chapter concentrates on developing linear discriminant function to separate the two groups under the status of mortality of children between 0-5 years of age using risk factors collected from the survey data as described in section 3.3. Leave-out method used to assess the performance of the final model using SPSS.

5.1 Development of Linear Discriminant Model

Linear discriminant analysis is performed well, when there are continuous variables functioning as the independent variables. In real scenario, both continuous and categorical variables are taking as the independent variables. As stated by Abdi (2007), discriminant correspondence analysis is used to test the categorical independent variables. Some literature proves that, when the discriminant analysis was found to be robust enough to the inclusion of categorical variables as independent variables, in such cases categorical variables can be converted to dummy variables. When there are combinations of both categorical and continuous variables for the independent variables, discriminant analysis can be carried out without any doubt. SPSS is automatically converting the categorical variables into dummy variables and performs the discriminant analysis.

5.1.1 Selection of Discriminating Variables

Stepwise discriminant analysis was carried out in the process of selecting variables with the significant level of F test of the covariance analysis (Alayande, 2015). According to the significance level of 0.05, all the independent variables were tested for retaining or leaving out in model development.

Wilks' lambda statistics was used to test the null hypothesis of the equal group means in testing the significance of discriminant function (Betz, 1987). In this scenario, the null hypothesis can be expressed as follows:

H_0: The two groups of status of mortality are significantly different with respect to each predictor variables

H_1: The two groups of status of mortality are not significantly different with respect to each predictor variables

Table 5.1: Significance of Discriminating Variables

Independent Variables	Wilks' Lambda	F	df1	df2	Sig.
Gender of Child (X1)	0.998	21.894	1	10587	0.000
Marital Status (X3)	0.986	153.187	1	10587	0.000
Mother's Literacy (X4)	0.972	306.618	1	10587	0.000
Status of Antenatal Care (X5)	0.824	2256.186	1	10587	0.000
Status of vitamin/medicine Usage (X6)	0.940	681.602	1	10587	0.000
Delivery Type (X7)	0.909	1065.668	1	10587	0.000
Pregnancy Duration (X8)	0.971	317.344	1	10587	0.000
Decision Making Ability (X9)	1.000	1.017	1	10587	0.313

The test of F statistics for Wilks's Lambda is significant for the seven variables (p-value < 0.05). This confirmed that the selected variables have a significant impact to differentiate the two groups of the status of mortality of children of age between 0-5 years.

For the analysis of the status of mortality, it was selected nine, most influential variables among 100s' of predictors as mentioned on the chapter 3 from the DHS survey. Finally, according to the significant level (p value < 0.05), only X1, X3-X9 (gender, marital status, mother's literacy and status of antenatal care, status of vitamin/medicine usage, delivery type and pregnancy duration) have been retained and the discriminant function was developed using the significant variables only.

The summary results of stepwise selection approach are shown in table 5.2.

Table 5.2: Results of Stepwise Method in DA

Step	Independent Variables	Tolerance	Min. Tolerance	Sig. of F to Enter	Wilks' Lambda
0	Gender of Child (X1)	1.000	1.000	0.000	0.998
	Marital Status (X3)	1.000	1.000	0.000	0.986
	Mother's Literacy (X4)	1.000	1.000	0.000	0.972
	Status of Antenatal Care (X5)	1.000	1.000	0.000	0.824
	Status of vitamin/medicine Usage (X6)	1.000	1.000	0.000	0.940
	Delivery Type (X7)	1.000	1.000	0.000	0.909
	Pregnancy Duration (X8)	1.000	1.000	0.000	0.971
	Decision Making Ability (X9)	1.000	1.000	0.313	1.000
1	Gender of Child (X1)	1.000	1.000	0.000	0.823
	Marital Status (X3)	1.000	1.000	0.000	0.815
	Mother's Literacy (X4)	0.999	0.999	0.000	0.808
	Status of vitamin/medicine Usage (X6)	0.697	0.697	0.968	0.824
	Delivery Type (X7)	0.958	0.958	0.000	0.791
	Pregnancy Duration (X8)	1.000	1.000	0.000	0.805
	Decision Making Ability (X9)	1.000	1.000	0.165	0.824
2	Gender of Child (X1)	1.000	0.958	0.000	0.789
	Marital Status (X3)	1.000	0.958	0.000	0.782
	Mother's Literacy (X4)	0.999	0.957	0.000	0.774
	Status of vitamin/medicine Usage (X6)	0.697	0.674	0.776	0.791
	Pregnancy Duration (X8)	0.998	0.956	0.000	0.775
	Decision Making Ability (X9)	1.000	0.958	0.164	0.790
3	Gender of Child (X1)	1.000	0.957	0.000	0.773
	Marital Status (X3)	0.995	0.957	0.000	0.768
	Status of vitamin/medicine Usage (X6)	0.690	0.674	0.234	0.774
	Pregnancy Duration (X8)	0.998	0.955	0.000	0.759
	Decision Making Ability (X9)	1.000	0.957	0.238	0.774
4	Gender of Child (X1)	1.000	0.955	0.000	0.758
	Marital Status (X3)	0.992	0.955	0.000	0.752
	Status of vitamin/medicine Usage (X6)	0.689	0.674	0.537	0.759
	Decision Making Ability (X9)	1.000	0.955	0.222	0.759
5	Gender of Child (X1)	1.000	0.955	0.000	0.751
	Status of vitamin/medicine Usage (X6)	0.689	0.674	0.507	0.752
	Decision Making Ability (X9)	0.921	0.914	0.000	0.751
6	Status of vitamin/medicine Usage (X6)	0.689	0.674	0.509	0.751
	Decision Making Ability (X9)	0.921	0.914	0.000	0.750

Table 5.2: Results of Stepwise Method in DA (Continued)

Step	Independent Variables	Tolerance	Min. Tolerance	Sig. of F to Enter	Wilks' Lambda
7	Status of vitamin/medicine Usage (X6)	0.689	0.673	0.479	0.750

According to the results depicted in table 5.2 and table 5.3, obtained from stepwise discriminant analysis, there is a statistically significant relationship between the predictor variables and mortality status except the predictor, status of vitamin/medicine usage.

Table 5.3: Variables not in the Analysis based on the DA

Step	Independent Variables	Tolerance	Min. Tolerance	Sig. of F to Enter	Wilks' Lambda
	Gender of Child (X1)	1.000	1.000	0.000	0.998
	Marital Status (X3)	1.000	1.000	0.000	0.986
	Mother's Literacy (X4)	1.000	1.000	0.000	0.972
	Status of Antenatal Care (X5)	1.000	1.000	0.000	0.824
	Status of vitamin/medicine Usage (X6)	1.000	1.000	0.000	0.940
	Delivery Type (X7)	1.000	1.000	0.000	0.909
	Pregnancy Duration (X8)	1.000	1.000	0.000	0.971
0	Decision Making Ability (X9)	1.000	1.000	0.313	1.000
	Gender of Child (X1)	1.000	1.000	0.000	0.823
	Marital Status (X3)	1.000	1.000	0.000	0.815
	Mother's Literacy (X4)	0.999	0.999	0.000	0.808
	Status of vitamin/medicine Usage (X6)	0.697	0.697	0.968	0.824
	Delivery Type (X7)	0.958	0.958	0.000	0.791
	Pregnancy Duration (X8)	1.000	1.000	0.000	0.805
1	Decision Making Ability (X9)	1.000	1.000	0.165	0.824
	Gender of Child (X1)	1.000	0.958	0.000	0.789
	Marital Status (X3)	1.000	0.958	0.000	0.782
	Mother's Literacy (X4)	0.999	0.957	0.000	0.774
	Status of vitamin/medicine Usage (X6)	0.697	0.674	0.776	0.791
	Pregnancy Duration (X8)	0.998	0.956	0.000	0.775
2	Decision Making Ability (X9)	1.000	0.958	0.164	0.790
	Gender of Child (X1)	1.000	0.957	0.000	0.773
	Marital Status (X3)	0.995	0.957	0.000	0.768
	Status of vitamin/medicine Usage (X6)	0.690	0.674	0.234	0.774
	Pregnancy Duration (X8)	0.998	0.955	0.000	0.759
3	Decision Making Ability (X9)	1.000	0.957	0.238	0.774

Table 5.3: Variables not in the Analysis based on the DA (Continued)

Step	Independent Variables	Tolerance	Min. Tolerance	Sig. of F to Enter	Wilks' Lambda
4	Gender of Child (X1)	1.000	0.955	0.000	0.758
	Marital Status (X3)	0.992	0.955	0.000	0.752
	Status of vitamin/medicine Usage (X6)	0.689	0.674	0.537	0.759
	Decision Making Ability (X9)	1.000	0.955	0.222	0.759
5	Gender of Child (X1)	1.000	0.955	0.000	0.751
	Status of vitamin/medicine Usage (X6)	0.689	0.674	0.507	0.752
	Decision Making Ability (X9)	0.921	0.914	0.000	0.751
6	Status of vitamin/medicine Usage (X6)	0.689	0.674	0.509	0.751
	Decision Making Ability (X9)	0.921	0.914	0.000	0.750
7	Status of vitamin/medicine Usage (X6)	0.689	0.673	0.479	0.750

5.2 Estimation of Canonical Discriminant Function

Table 5.4: Unstandardized Coefficients of Discriminant Function

Independent Variables	Function (1)
Gender of Child (X1)	-0.164
Marital Status (X3)	-0.605
Mother's Literacy (X4)	0.524
Status of Antenatal Care (X5)	1.747
Delivery Type (X7)	0.842
Pregnancy Duration (X8)	0.582
Decision Making Ability (X9)	0.076
(Constant)	-7.955

Unstandardized coefficients

Unstandardized coefficients of canonical discriminant function are used to compute the discriminant scores by summation of produced values with predictors plus the constant. Based on the results in table 5.4, the discriminant function can be written as,

$$FD = -7.955 - 0.164 * X1 - 0.605 * X3 + 0.524 * X4 + 1.747 * X5 + 0.842 * X7 + 0.582 * X8 + 0.076 * X9 \quad (5.1)$$

It is observed (eq. 5.1) that gender(X1) and marital status(X3) define the negative relationship with mortality status while mother’s literacy(X4), pregnancy duration(X8) and decision-making ability(X9), status of antenatal care(X5) and delivery type(X7) are positively attached with the two cases of the child mortality.

In order to test the multicollinearity of the variables in the model, the correlation matrix was obtained for categorical variables using Spearman's rank order correlation. The continuous variable X8, pregnancy duration was also converted into categorical variable for three levels which were 'less than 6 months', '7-8 months' and 'more than 9 months'.

Table 5.5: Correlation Matrix of the Independent Variables

	Gender of Child	Marital Status	Mother's Literacy	Status of Antenatal Care	Vitamin/medicine Usage	Delivery Type	Decision Making Ability	Pregnancy Duration
Gender of Child	1.000							
Marital Status	0.006	1.000						
	0.532							
Mother's Literacy	0.011	-.068**	1.000					
	0.239	0.000						
Status of Antenatal Care	-.021*	-.044**	.099**	1.000				
	0.032	0.000	0.000					
Status of vitamin/medicine Usage	0.011	0.018	-.143**	-.588**	1.000			
	0.277	0.059	0.000	0.000				
Delivery Type	-0.012	-.035**	.031**	.304**	-.167**	1.000		
	0.219	0.000	0.001	0.000	0.000			
Decision Making Ability	-0.004	.187**	0.017	-0.006	-0.013	0.002	1.000	
	0.644	0.000	0.088	0.559	0.173	0.851		
Pregnancy Duration	-0.014	0.010	0.009	.088**	-.051**	.114**	0.005	1.000
	0.157	0.299	0.351	0.000	0.000	0.000	0.594	

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Based on the correlation matrix (Table 5.5), it can be concluded that almost all the correlation coefficient values are very low (< 0.1), even though they are statistically significant due to high number of observation (n = 10,589). Hence, this is a good indication to claim that there is no significant multicollinearity problem among the selected variables.

The lowest significant correlation (r = -0.588, p = 0.000) was found between status of antenatal care and status of vitamin/medicine usage. Status of antenatal care and delivery type is highly correlated significantly (r = 0.304, p = 0.000). Decision making ability is also relatively highly correlated (r = 0.187, p = 0.000) with marital status. When it is considered the marital status and mother's literacy, significant correlation (r = -0.068, p- = 0.000) can also be seen with each other.

Figure 5.1 demonstrates the distribution of canonical discriminant scores resulted from the equation 5.1 to the predicted group membership of status of mortality. It is a graphical representation of classification capacity of the final model, two classes of two different mean values. Canonical discriminant scores are varied from -5.491 (minimum value) to 2.242 (maximum value) with the mean value of 0. Allocated groups are derived using mean value taking as the cutoff value.

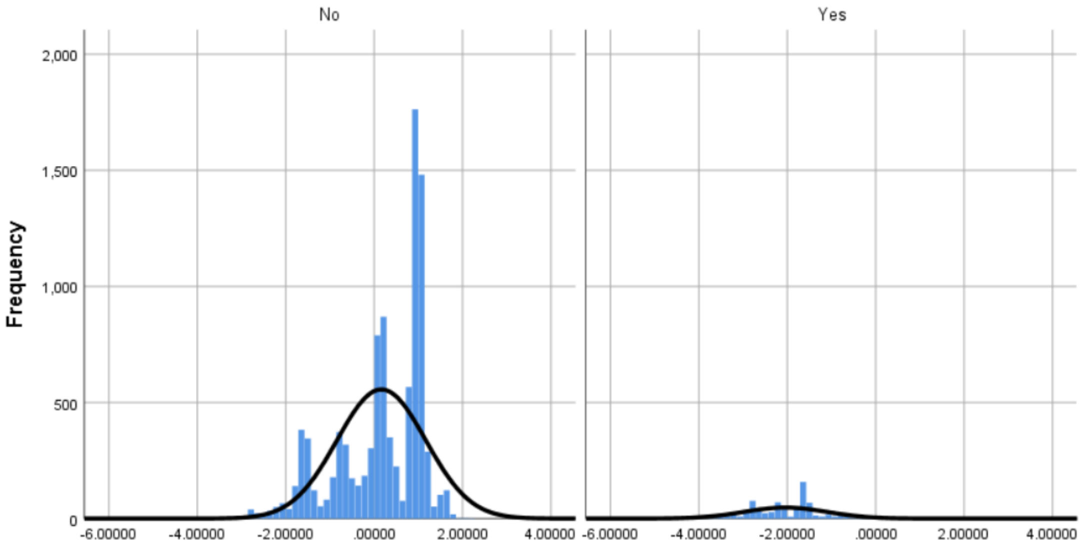


Figure 5.1: Distribution of the Canonical Discriminant Scores

Relative strength of discriminant variables is illustrated with constructing the structural canonical coefficients, while distinctive contribution of predictors to the discriminant function by using the standardized canonical coefficients as shown in table 5.6.

Table 5.6: Coefficients of Canonical Function

Independent Variables	Standardized (1)	Structure (1)
Gender of Child (X1)	-0.082	-0.079
Marital Status (X3)	-0.219	-0.208
Mother's Literacy (X4)	0.268	0.295
Status of Antenatal Care (X5)	0.708	0.799
Delivery Type (X7)	0.394	0.549
Pregnancy Duration (X8)	0.288	0.300
Decision Making Ability (X9)	0.084	0.017
Status of vitamin/medicine Usage (X6)	0.000	-0.451

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

It is indicated a greater potential in discrimination When the higher value of standardized canonical coefficient exists. Standardized canonical coefficients are differed from the range between -0.279 to 0.708. Status of vitamin/medicine usage shows the zero value, that indicated the non-existence of that predictor in modelling the child mortality. Delivery type and status of antenatal care act as the most influential discriminating variable, while gender of child and marital status act as least influential discriminating variable.

5.3 Estimation of Fisher's Linear Discriminant Function

Table 5.7: Coefficients of FLD in Stepwise DA

Independent Variables	Mortality Status	
	No(1)	Yes (2)
Gender of Child (X1)	1.860	2.221
Marital Status (X3)	5.095	6.431
Mother's Literacy (X4)	11.170	10.014
Status of Antenatal Care (X5)	3.958	0.101
Delivery Type (X7)	0.583	-1.276
Pregnancy Duration (X8)	36.618	35.333
Decision Making Ability (X9)	1.951	1.784
(Constant)	-189.090	-173.603

Fisher's linear discriminant functions

For the comparison of two groups related to the mortality status, two classification scores were generated for each observation of two classification functions using fisher's linear discriminant functioning. Then the groups were allocated based on the higher scores of the two discriminant functions. table 5.7 illustrates the coefficients of the two linear discriminant functions.

According to the results in table 5.7, two linear discriminant functions were derived as follows:

$$F(LDF1) = -189.09 + 1.86 * X1 + 5.095 * X3 + 11.17 * X4 + 3.958 * X5 + 0.583 * X7 + 36.618 * X8 + 1.951 * X9 \quad (5.2)$$

$$F(LDF2) = -173.603 + 2.221 * X1 + 6.431 * X3 + 10.014 * X4 + 0.101 * X5 - 1.276 * X7 + 35.333 * X8 + 1.784 * X9 \quad (5.3)$$

Figure 5.2 and figure 5.3 demonstrate the distribution of fisher's linear discriminant scores resulted from the equation 5.2 and equation 5.3 respectively to the predicted group membership of status of mortality.

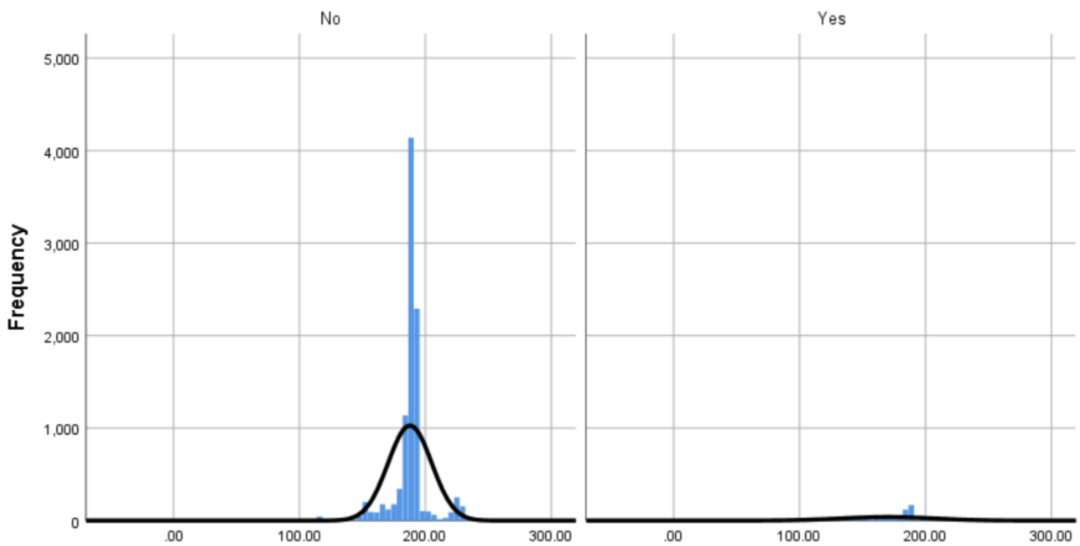


Figure 5.2: Distribution of the Fisher's Linear Discriminant Scores for LDF1

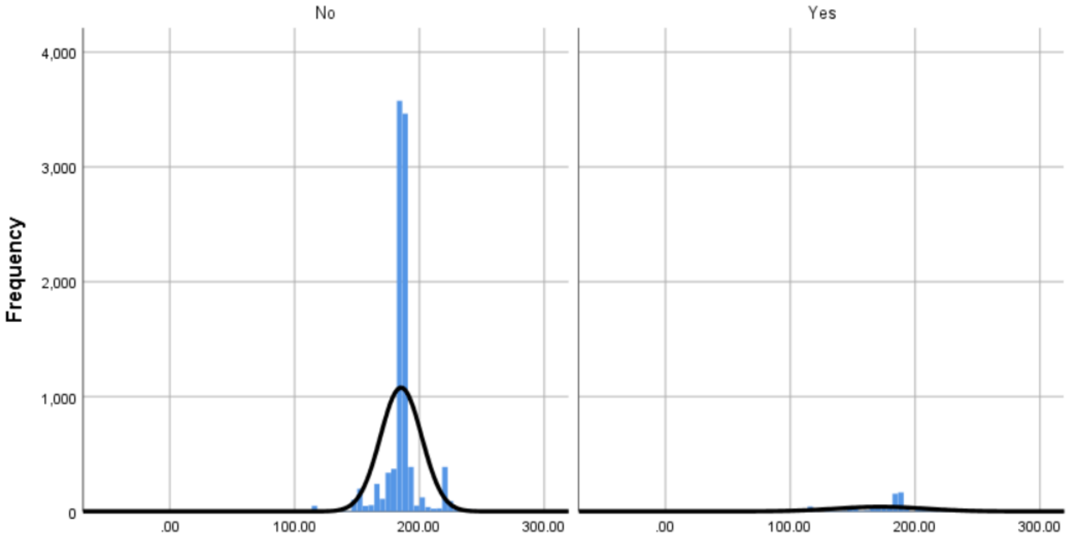


Figure 5.3: Distribution of the Fisher's Linear Discriminant Scores for LDF2

Eqn. 5.1 of canonical discriminant function and eqn.5.2 & eqn.5.3 of fisher's discriminant functions were used to compute the discriminant scores for the allocation of mortality status into two groups. Both were derived with the same statistical results of outputs. Hence canonical discriminant scores were selected for the further analysis of the child mortality.

5.4 Validation of the Developed Model

In this section, the following hypothesis are tested:

H_0: The two groups of the status of mortality are significantly associated with predictor variables

H_1: The two groups of the status of mortality are not significantly associated with predictor variables

The statistics, Wilk's lambda is used to test the significance of the model in discriminant functioning. The canonical function was significant as the test statistic, p-value < 0.05. Wilks` lambda which is distributed under chi-square is significant (Table 5.8). Thus, the developed discriminant function can be used for demonstrating the group membership of child mortality.

Table 5.8: Significance Test of Predictors in Canonical Function

Function	Eigenvalue	Canonical Correlation	Wilks' Lambda	Chi-square	Df	Sig.
1	0.334a	0.500	0.750	3049.25	7	0.000

a First 1 canonical discriminant functions were used in the analysis.

5.5 Predictive Power of the Model

The results of the classification of all the observations based on the above linear discriminant function is summarized in table 5.9.

Table 5.9: Classification Results of the Discriminant Function

		Mortality Status	Predicted Group Membership		Total
			No	Yes	
Original	Count	No	8272	1534	9806
		Yes	65	718	783
	%	No	84.4	15.6	100
		Yes	8.3	91.7	100

a 84.9% of original grouped cases correctly classified.

It is stated that 8272 of 9806, the survival group were correctly classified and 718 of 783, the death group were correctly classified from child births. Finally, it was able to classify 8990 out of 10589 child births correctly. According to the Table 5.8, It has been correctly classified the 84.9% (ACCR) of grouped cases. Then the final model has a great potential on classification of mortality status. Then Apparent Error Rate (AER) can be claimed as the 15.1%.

5.6 Summary of Chapter 5

Modelling the child mortality in Sri Lanka with discriminant analysis approach, it was evidenced that of the selected variables, gender, marital status, mother's literacy, status of antenatal care, delivery type and pregnant duration were found as the significantly influential variables, while status of vitamin/medicine usage was not identified as a significant variable. The percentage of overall correctly classification by the developed linear discriminant function is 84.9% irrespective of the status of mortality.

CHAPTER 6

DEVELOPMENT OF BINARY LOGISTIC REGRESSION MODEL

This chapter provides the second phase of model building via binary logistic regression approach. The stepwise binary logistic regression was carried out in modelling the child mortality of age 0-5 years of children.

6.1 Binary Logistic Model with all Eight Variables

For the development of binary logistic regression model, initially it has been selected 8 variables, X1, X3-X9 (gender, marital status, mother's literacy, status of antenatal care, status of vitamin/medicine usage, delivery type and pregnancy duration and decision-making ability). In this scenario, the following hypothesis were tested:

H_0: The log odds of the status of child mortality are not significantly associated with independent variables

H_1: The log odds of the status of child mortality are significantly associated with independent variables

Of the variables: X1, X3-X9, all variables except X6 which was not significantly associated with the status of the mortality were included simultaneously for the model and corresponding SPSS outputs are shown in tables 6.1.

Table 6.1: Omnibus Test of Model Coefficients

		Chi-square	df	Sig.
Step 7	Step	13.041	1	0.000
	Block	2548.381	14	0.000
	Model	2548.381	14	0.000

The Omnibus tests of model (Table 6.1) indicates that a test of the null hypothesis that adding the all eight variables to the model has not significantly increased is rejected as the corresponding chi-square statistics is significant (x_8^2 =2548.381, p-value = 0.000). Thus, the model with seven variables can be accepted as significant variable to explain the response variables of all the subjects.

Table 6.2: Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
7	3036.789[b]	0.214	0.522

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.
b. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

The -2log (likelihood) statistic in table 6.2 is also confirmed the statistical significance of the overall model with seven variables. The value of Cox & Snell R Square and Nagelkerke R Square statistics indicate that fitted logistic model is been able to explain only 21.4%   is a one of the forms of Cox & Snell coefficient modifications. Those values indicate that the amount of variance that can be explained by the model vary between 21.4% and 52.2%.   However, it should be pointed out that these statistics are not powerful as $\chi 2$ in multiple regression.

Table 6.3: Hosmer & Lemeshow Test Statistics

Step	Chi-square	df	Sig.
7	41.058	7	0.000

Hosmer and Lemeshow test statistic is used to compare the fitted expected values and actual values by groups (Memic, 2015). As stated in Table 6.3, Hosmer and Lemeshow test has been produced chi-square test statistic is significant ($x_8^2 = 41.058, p = 0.000$) confirming that the fitted model is significantly better than the model with intercept only.

The details of coefficients of the fitted model is shown in table 6.4.

Table 6.4: Properties of the Coefficients of the fitted model

Step	Independent Variables	B	S.E.	Wald	df	Sig.
	Child Gender (X1)(1)	-0.339	0.094	12.931	1	0.000
	Marital Status (X3)			70.859	2	0.000
	Marital Status (X3)(1)	-2.133	0.256	69.532	1	0.000
	Marital Status (X3)(2)	-2.369	0.383	38.176	1	0.000
	Mother's Literacy (X4)			124.680	4	0.000
	Mother's Literacy (X4)(1)	0.078	0.513	0.023	1	0.879
	Mother's Literacy (X4)(2)	0.067	0.513	0.017	1	0.896
	Mother's Literacy (X4)(3)	-1.167	0.497	5.514	1	0.019
	Mother's Literacy (X4)(4)	-18.858	8229.049	0.000	1	0.998
	Status of Antenatal Care (X5)(1)	3.359	0.152	485.477	1	0.000
	Delivery Type (X7)(1)	2.426	0.149	264.587	1	0.000
	Pregnancy Duration (X8)	-0.701	0.067	110.914	1	0.000
	Decision Making Ability (X9)			31.911	4	0.000
	Decision Making Ability (X9)(1)	1.163	0.208	31.300	1	0.000
	Decision Making Ability (X9)(2)	0.763	0.207	13.525	1	0.000
	Decision Making Ability (X9)(3)	0.787	0.177	19.758	1	0.000
	Decision Making Ability (X9)(4)	1.098	0.496	4.898	1	0.027
7	Constant	2.204	0.821	7.210	1	0.007

Results in table 6.4 very clearly indicate that all the variables are significant on the response variable, at least 5% level.

The model can be written as:

$$\text{Logit (p)} = \text{Log}\left(\frac{p}{1-p}\right) \qquad (6.1)$$

$$= -0.339 * (x_{1=1}) - 2.133 * (x_{3=1}) -2.369 * (x_{3=2}) + 0.078 * (x_{4=1}) + 0.067 * (x_{4=2})$$

$$– 1.167 * (x_{4=3}) – 18.858 * (x_{4=4}) + 3.359 * (x_{5=1}) + 2.426 * (x_{7=1}) – 0.701 * (x_{8})$$

$$+ 1.163 * (x_{9=1}) + 0.763 * (x_{9=2}) + 0.787 * (x_{9=3}) + 1.098 * (x_{9=4}) + 2.204 \qquad (6.2)$$

6.2 Development of Model via Stepwise Forward (LR-Wald/Likelihood)

Table 6.5: Results of Hosmer and Lemeshow Test under Stepwise Forward (LR-Wald/Likelihood)

Step	Chi-square	df	Sig.
1	0.000	0	0.000
2	42.868	2	0.000
3	38.567	4	0.000
4	66.527	5	0.000
5	54.086	6	0.000
6	25.823	7	0.001
7	41.058	7	0.000

By including all the significant variables, a model was developed via Stepwise Forward (Wald) and Stepwise Forward (Likelihood) methods using entry probability of 0.05 and removal probability of 0.1. The important outputs are shown in Tables 6.5-6.6.

Table 6.6: Variables in the Model Obtained in Stepwise Forward (Wald/Likelihood) Method

Step	Independent Variables	B	S.E.	Wald	df	Sig.
1	Status of Antenatal Care (X5)(1)	3.942	0.148	707.792	1	0.000
	Constant	-5.028	0.142	1255.986	1	0.000
2	Status of Antenatal Care (X5)(1)	3.473	0.150	536.543	1	0.000
	Delivery Type (X7)(1)	2.448	0.143	292.551	1	0.000
	Constant	-6.474	0.186	1212.480	1	0.000
3	Mother's Literacy (X4)			153.798	4	0.000
	Mother's Literacy (X4)(1)	0.155	0.506	0.094	1	0.759
	Mother's Literacy (X4)(2)	-0.039	0.506	0.006	1	0.939
	Mother's Literacy (X4)(3)	-1.248	0.491	6.468	1	0.011
	Mother's Literacy (X4)(4)	-19.299	8052.166	0.000	1	0.998
	Status of Antenatal Care (X5)(1)	3.386	0.151	503.975	1	0.000
	Delivery Type (X7)(1)	2.455	0.145	286.289	1	0.000
	Constant	-5.434	0.525	107.284	1	0.000
4	Mother's Literacy (X4)			157.161	4	0.000
	Mother's Literacy (X4)(1)	0.016	0.514	0.001	1	0.975
	Mother's Literacy (X4)(2)	-0.134	0.515	0.068	1	0.794
	Mother's Literacy (X4)(3)	-1.404	0.499	7.900	1	0.005

Table 6.6: Variables in the Model Obtained in Stepwise Forward (Wald/Likelihood) Method (Continued)

Step	Independent Variables	B	S.E.	Wald	df	Sig.
	Mother's Literacy (X4)(4)	-19.349	8060.866	0.000	1	0.998
	Status of Antenatal Care (X5)(1)	3.363	0.151	493.279	1	0.000
	Delivery Type (X7)(1)	2.463	0.148	277.068	1	0.000
	Pregnancy Duration (X8)	-0.650	0.066	97.532	1	0.000
4	Constant	0.475	0.792	0.360	1	0.549
	Marital Status (X3)			49.473	2	0.000
	Marital Status (X3)(1)	-1.366	0.198	47.691	1	0.000
	Marital Status (X3)(2)	-1.677	0.347	23.304	1	0.000
	Mother's Literacy (X4)			131.261	4	0.000
	Mother's Literacy (X4)(1)	0.017	0.511	0.001	1	0.973
	Mother's Literacy (X4)(2)	-0.046	0.511	0.008	1	0.929
	Mother's Literacy (X4)(3)	-1.267	0.496	6.539	1	0.011
	Mother's Literacy (X4)(4)	-19.164	8077.584	0.000	1	0.998
	Status of Antenatal Care (X5)(1)	3.353	0.152	486.132	1	0.000
	Delivery Type (X7)(1)	2.429	0.149	266.835	1	0.000
	Pregnancy Duration (X8)	-0.683	0.066	105.820	1	0.000
5	Constant	1.994	0.819	5.932	1	0.015
	Marital Status (X3)			72.343	2	0.000
	Marital Status (X3)(1)	-2.147	0.255	70.943	1	0.000
	Marital Status (X3)(2)	-2.397	0.384	39.056	1	0.000
	Mother's Literacy (X4)			124.779	4	0.000
	Mother's Literacy (X4)(1)	0.065	0.513	0.016	1	0.899
	Mother's Literacy (X4)(2)	0.052	0.513	0.010	1	0.919
	Mother's Literacy (X4)(3)	-1.176	0.498	5.585	1	0.018
	Mother's Literacy (X4)(4)	-18.895	8196.405	0.000	1	0.998
	Status of Antenatal Care (X5)(1)	3.362	0.152	486.854	1	0.000
	Delivery Type (X7)(1)	2.433	0.149	266.327	1	0.000
	Pregnancy Duration (X8)	-0.704	0.067	111.464	1	0.000
	Decision Making Ability (X9)			32.625	4	0.000
	Decision Making Ability (X9)(1)	1.170	0.207	31.837	1	0.000
	Decision Making Ability (X9)(2)	0.763	0.207	13.604	1	0.000
	Decision Making Ability (X9)(3)	0.799	0.177	20.436	1	0.000
	Decision Making Ability (X9)(4)	1.143	0.495	5.330	1	0.021
6	Constant	2.087	0.820	6.480	1	0.011

Table 6.6: Variables in the Model Obtained in Stepwise Forward (Wald/Likelihood) Method (Continued)

Step	Independent Variables	B	S.E.	Wald	df	Sig.
	Child Gender (X1)(1)	-0.339	0.094	12.931	1	0.000
	Marital Status (X3)			70.859	2	0.000
	Marital Status (X3)(1)	-2.133	0.256	69.532	1	0.000
	Marital Status (X3)(2)	-2.369	0.383	38.176	1	0.000
	Mother's Literacy (X4)			124.680	4	0.000
	Mother's Literacy (X4)(1)	0.078	0.513	0.023	1	0.879
	Mother's Literacy (X4)(2)	0.067	0.513	0.017	1	0.896
	Mother's Literacy (X4)(3)	-1.167	0.497	5.514	1	0.019
	Mother's Literacy (X4)(4)	-18.858	8229.049	0.000	1	0.998
	Status of Antenatal Care (X5)(1)	3.359	0.152	485.477	1	0.000
	Delivery Type (X7)(1)	2.426	0.149	264.587	1	0.000
	Pregnancy Duration (X8)	-0.701	0.067	110.914	1	0.000
	Decision Making Ability (X9)			31.911	4	0.000
	Decision Making Ability (X9)(1)	1.163	0.208	31.300	1	0.000
	Decision Making Ability (X9)(2)	0.763	0.207	13.525	1	0.000
	Decision Making Ability (X9)(3)	0.787	0.177	19.758	1	0.000
	Decision Making Ability (X9)(4)	1.098	0.496	4.898	1	0.027
7	Constant	2.204	0.821	7.210	1	0.007

The chi-square statistics at step 7 is not significant at 95% confident interval, it can be concluded that the observed data and fitted data using the model is not significantly different and model 7 can be accepted. It can be seen the results in table 6.4 and table 6.6 are the same.

As explained in section 6.2, a model was developed using stepwise forward (likelihood) and the corresponding results are shown in table 6.6.

Identified models in both wald and likelihood methods, under stepwise selection are the same, thus it is confirming the results are invariant by the criteria of forward stepwise.

6.3 Development of the model via Stepwise Backward (LR-Likelihood)

As explained above a model was also developed using Backwards Stepwise (likelihood) in order to test the validity of model irrespective of the methodology. The two important Tables with respect to backward selection method are shown in table 6.7 & table 6.8.

Table 6.7: Results of Hosmer and Lemeshow Test under Stepwise Backward (Likelihood)

Step	Chi-square	df	Sig.
1	0.000	0	0.000
2	42.868	2	0.000
3	38.567	4	0.000
4	66.527	5	0.000
5	54.086	6	0.000
6	25.823	7	0.001
7	41.058	7	0.000

Table 6.8: Properties of the Coefficients of the Model Developed via Stepwise Backward (Likelihood)

Step	Independent Variables	B	S.E.	Wald	df	Sig.
	Child Gender (X1)(1)	-0.340	0.094	12.949	1	0.000
	Marital Status (X3)			70.883	2	0.000
	Marital Status (X3)(1)	-2.137	0.256	69.531	1	0.000
	Marital Status (X3)(2)	-2.378	0.384	38.399	1	0.000
	Mother's Literacy (X4)			125.101	4	0.000
	Mother's Literacy (X4)(1)	0.085	0.512	0.028	1	0.868
	Mother's Literacy (X4)(2)	0.076	0.513	0.022	1	0.882
	Mother's Literacy (X4)(3)	-1.166	0.496	5.513	1	0.019
	Mother's Literacy (X4)(4)	-18.862	8236.795	0.000	1	0.998
	Status of Antenatal Care (X5)(1)	3.475	0.253	188.742	1	0.000
	Status of vitamin/medicine Usage (X6)			0.511	2	0.774
	Status of vitamin/medicine Usage (X6)(1)	0.218	0.305	0.511	1	0.475
1	Status of vitamin/medicine Usage (X6)(2)	-15.930	2841.855	0.000	1	0.996

Table 6.8: Properties of the Coefficients of the Model Developed via Stepwise Backward (Likelihood) (Continued)

Step	Independent Variables	B	S.E.	Wald	df	Sig.
	Delivery Type (X7)(1)	2.426	0.149	264.410	1	0.000
	Pregnancy Duration (X8)	-0.701	0.067	110.760	1	0.000
	Decision Making Ability (X9)			31.918	4	0.000
	Decision Making Ability (X9)(1)	1.163	0.208	31.313	1	0.000
	Decision Making Ability (X9)(2)	0.763	0.207	13.513	1	0.000
	Decision Making Ability (X9)(3)	0.787	0.177	19.758	1	0.000
	Decision Making Ability (X9)(4)	1.096	0.496	4.884	1	0.027
1	Constant	2.093	0.843	6.171	1	0.013

According to the results shown in tables 6.4, 6.6, and 6.8. It can be concluded that the same variables were identified as significantly influential variables on the response variable at 5% level. It confirms that the results are invariant on the method as well as criteria selected in variable selection.

There is a statistically significant impact from the predictor variables such as gender, marital status, mother's literacy, status of antenatal care, pregnancy duration, decision making ability and delivery type on the dependent variable, status of mortality.

6.4 Predicted Probability of Logistic Function

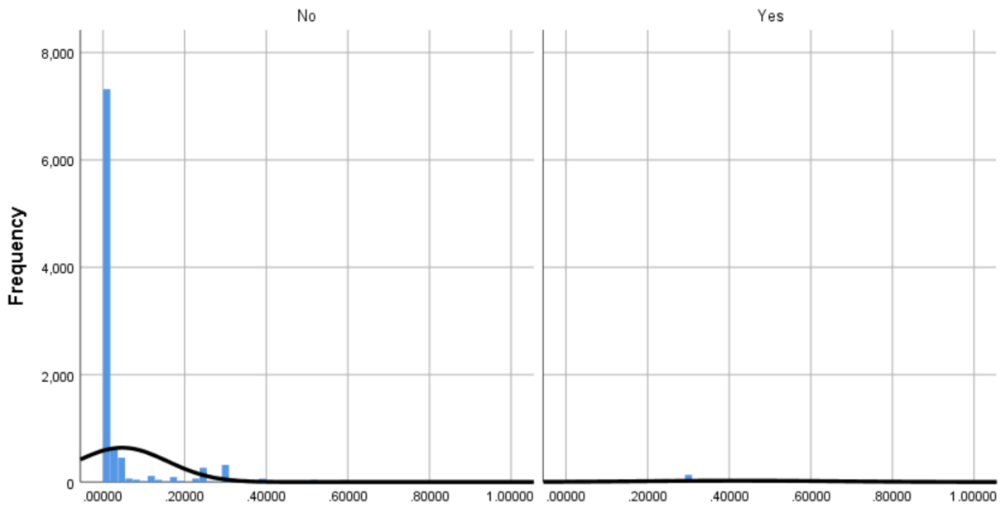


Figure 6.1: Distribution of Predicted Probability among Two Groups

Figure 6.1 illustrates the distribution of predicted probability generated from the equation 6.2 for the allocation of the group membership of the status of child mortality. It is a one of the graphical representations of classification capacity of the final model.

6.5 Predictive Power of the Model

As Table 6.9 illustrates that 9652 of 10006 were correctly classified from the survival group of child births, and 314 of 783 from the death group of child births were correctly classified by the analysis. Then it is concluded that the 9966 cases of child births out of 10,589 cases were correctly classified by the stepwise binary logistic model. Then it can be expressed that ACCR indicates the 94.1% and AER indicates the 5.9%. The final model was evaluated with a classification cut off point of 0.50.

Table 6.9: Classification Results of Logistic Regression Model

			Predicted Group Membership		
			Mortality Status		Percentage Correct
Observed			No	Yes	
Step 7	**Mortality Status**	No	9652	154	98.4
		Yes	469	314	40.1
	Overall Percentage				94.1

a. The cut value is .500

6.6 Summary of Chapter 6

Binary logistic model too identified that the variables: gender, marital status, mother's literacy, status of antenatal care, delivery type and pregnancy duration were found as the significantly influential variables, while status of vitamin/medicine usage was not significant. The identified model is invariant of type of method used.

Chapter 7

COMPARISON OF TWO MODELS

Linear discriminant analysis and binary logistic regression were used in addressing same research problem. This chapter provides the comparative approach in statistical analysis using misclassification, ROC curves, sensitivity/specificity and sample size variations.

7.1 Identification of Significant Variables

Modelling the child mortality, it was chosen the stepwise approach as the execution method of two statistical data models. Table 7.1 summarizes, how predictors shifted within the process of analysis.

Table 7.1: Summary Steps in Predictors Added/Removed

Step	Discriminant Function		Logistic Regression	
	IN	OUT	IN	OUT
1	Status of Antenatal Care (X5)		Status of Antenatal Care (X5)	
2	Delivery Type (X7)		Delivery Type (X7)	
3	Mother's Literacy (X4)		Mother's Literacy (X4)	
4	Pregnancy Duration (X8)		Pregnancy Duration (X8)	
5	Marital Status (X3)		Marital Status (X3)	
6	Gender of Child (X1)		Gender of Child (X1)	
7	Decision Making Ability (X9)	Status of vitamin/medicine Usage (X6)	Decision Making Ability (X9)	Status of vitamin/medicine Usage (X6)

Table 7.1 provides the evidence of how predictor variables behaved, when the process of selecting the variables into the final model at each step according to their significant level. In discriminating the status of child mortality, it was strongly evidenced that both technologies resulted with formulating the final model by selecting the almost similar input variables such as, X1, X3 - X5 and X7 – X9. They are namely, gender(X1), marital status(X3), mother’s literacy(X4), status of antenatal care(X5), delivery type(X7), pregnancy duration(X8) and decision-making ability(X9). Status of vitamin/medicine usage(X6) was not statistically significant under both models.

7.2 Probability of Misclassification

As the first step, it was considered two statistical criteria: Apparent Error Rate (AER) and Apparent Correct Classification Rate (ACCR) for the comparison of two statistical models (Section 5.5 and Section 6.5). The results of discriminant analysis yield that 84.9% of correctly classified original cases, while logistic regression obtained the 94.1% of correctly classified original cases to their respective groups of child mortality status.

For the better representation of all four conditions tested in two statistical methodologies, the summary table of the rate of type I error (misclassification of survival to death), rate of type II error (misclassification of death to survival) and rate of correct classification demonstrates in table 7.2 below:

Table 7.2: Performance Summary with Misclassifications

	Error Classification (%)		Correct Classification (%)
	Type I (%)	Type II (%)	
Discriminant Analysis	14.49	0.61	84.9
Logistic Regression	1.45	4.43	94.1

From the two statistical approach, it was found 14.49%, type I error and 0.61% type II error for discriminant function. Then considering the logistic regression, it was cleared that 1.45% for type I error and 4.43% for type II error in error classification. Furthermore, logistic regression has the higher rate in correct classification compared to the discriminate function.

The type II error is remarkably low in discriminant analysis, which was less than 1% in modelling child mortality, while type I error exhibited very high value compared to all the four values of error classification results. Then it is concluded that the overall performance is best in logistic regression approach, when it was produced low values in error classification compared to the values produced by discriminant analysis.

7.3 Receiver Operating Characteristic (ROC) Curve

Receiver Operating Characteristic (ROC) curve was also used to compare the discriminant function and logistic regression by graphing the status of child mortality against the probability of distribution of each model. Area under the curve (AUC) is the criteria for measuring the accuracy of the final model.

Table 7.3 indicates the test results of **a**rea under the ROC curve with the attributes of sample sizes, size of the area, standard error, 95% confidence interval in upper and lower bounds and significance tests for discriminant analysis and logistic regression models. Prediction of classification accuracy were evaluated with two different sample sizes, which were respectively the total sample size and the 5% of total sample size.

Table 7.3: Test Results of Area Under Curve

Sample Size	Analysis Method	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
					Lower Bound	Upper Bound
Total (N=10,589)	DA	0.931	0.005	0.000	0.922	0.940
	LR	0.938	0.004	0.000	0.930	0.947
5% of Total (n=573)	DA	0.959	0.009	0.000	0.941	0.977
	LR	0.893	0.015	0.000	0.864	0.923

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Considering the total sample size of each model building, the area under the ROC curve for the logistic regression model is 0.938, with 95% confidence interval 0.930-0.947, stand. Error 0.004 and p-value < 0.001 and for the discriminant functioning model, the area under the ROC curve is 0.931 with a 95% CI of 0.922-0.940, std. error 0.005 and p-value <0.001.

Beside the total sample size, when sample size equals to 573 of each model building, the area under the ROC curve for the logistic regression model is 0.893, with 95% confidence interval 0.864-0.923, stand. Error 0.015 and p-value < 0.001 and for the discriminant functioning model, the area under the ROC curve is 0.959 with a 95% CI of 0.941-0.977, std. error 0.009 and p-value <0.001.

Figures 7.1 illustrates the ROC curve for discriminant analysis model and Figure 7.2 presented the ROC curves for logistic regression model. Total sample size of 10,589 individuals were considered, when graphing both curves. For the effective assessment of modelling data at each cut-off points, ROC curves were potted for sensitivity (number of true positive) and 1 - specificity (number of true negative) rather than taking general specificity.

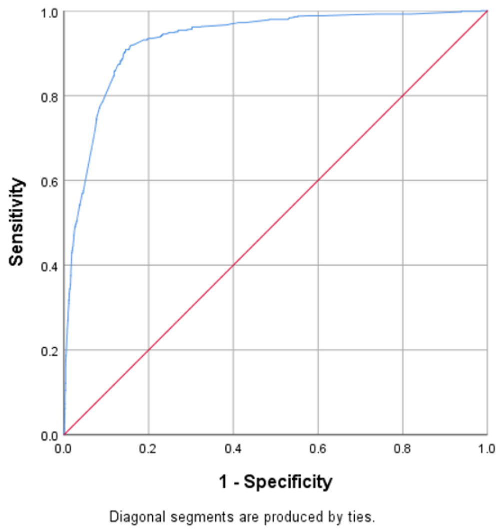


Figure 7.1: ROC Curve for Discriminant Function

Each Points on the ROC curve are the probability levels, which are resulted from model building and represents the sensitivity and 100% minus specificity (false positive rate) at different cut-off points at 95% confidence interval. (Montgomery et al.,1986) Changes of each cutoff points are exhibited clearly.

X-axis shows specificity, which means the probability that the models will correctly classify the mortality status of the default incorrectly classified with total default group. Then the Y-axis shows the sensitivity, which means the probability that status of mortality will correctly classify defaulted. The best model with powerful discriminating potential will be selected based on the sensitivity and specificity, which are obtained higher values (Gurny et al., 2013).

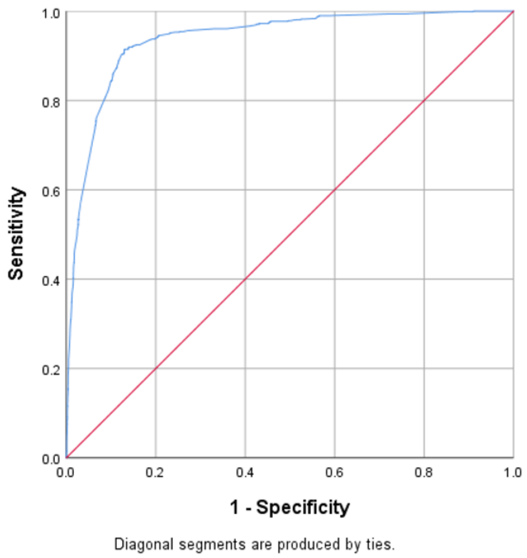


Figure 7.2: ROC Curve for Logistic Regression Model

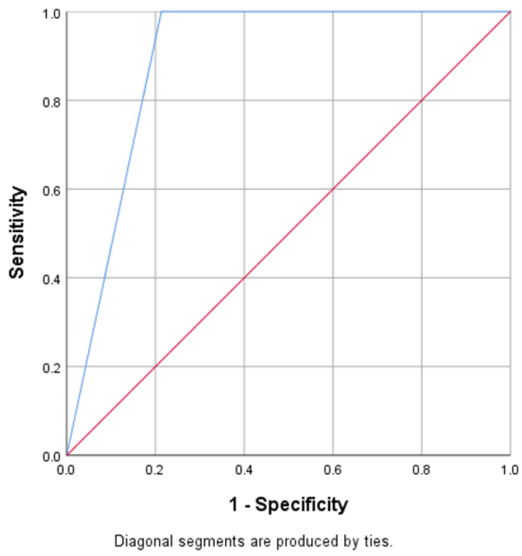
According to the graphical representation of figure 7.1 and figure 7.2, it is cleared that both models are covered almost same area sizes and differences between them can be neglected. The two curves are overlapped between them. The results are interpreted as the one of the methods of estimating the model performance which is indicating the performances of the two models are above the diagonal line, ensures best fit of models. The following hypothesis were tested.

H_0: AUC = 0.5

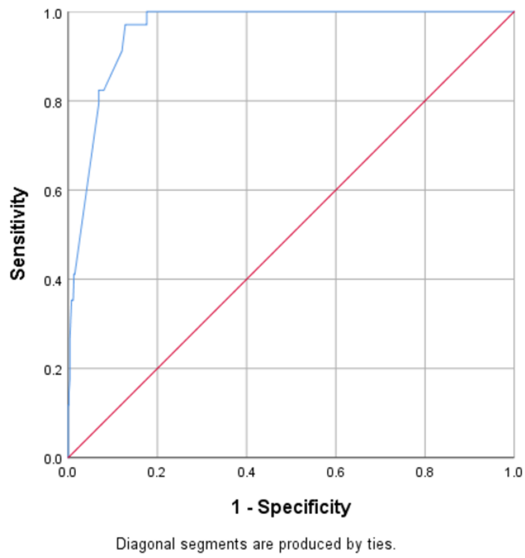
H_1: AUC $\neq$ 0.5

As evidence provided by the table 7.3, the null hypothesis was accepted, hence the value of area under the curve is significantly different from the value, 0.5.

Figures 7.3 displays the ROC curves for discriminant function and logistic regression using sub sample set of 5 percent from the original set of data, which are included 573 individuals of children. It is evidenced that both curves are representing the similar values of areas under curve and the difference between them are almost negligible.



Logistic Regression (n=573)



Discriminant Analysis (n=573)

Figure 7.3: Comparison of ROC Curves for Sample Size, n=5% of Total

In the classification function, when comparing the above two aforementioned models, logistic regression exceeds discriminant function a bit. According to the table 7.3, results of area under curve of 95% CI, the differences between areas under curve were non-considerable, and then no discriminating difference between the two models irrespective of the sample sizes.

7.4 Analysis of Sensitivity/ Specificity

In this section sensitivity and specificity were used to compare both models with the similar set of independent variables for the status of mortality as the dependent variable at each cut-off points. Cut-off probabilities were varied from the range of 0.1 to 1.0 at the increment of 0.1. The results in table 7.4 illustrates the sensitivity, specificity of both statistical approaches at optimal cutoff point for the status of child mortality.

Further logistic regression resulted with slightly lower values for sensitivity and slightly higher values for specificity, compared to the discriminant function.

Table 7.4: Analysis of Sensitivity and Specificity

	Sensitivity (%)	Specificity (%)
Discriminant Analysis	91.69	84.36
Logistic Regression	40.10	98.43

For the logistic regression model, at the correct classification rate (94.1%) sensitivity and specificity were respectively 40.1% and 94.43%. For the discriminant function, at the correct classification rate, sensitivity and specificity were respectively 91.69% and 84.36%.

7.5 Comparison of Accuracy with Sample Size Variations

The resulted models, which were used for testing the classification accuracy were executed with five different sample sizes from the original dataset for the purpose of estimating how sample sizes are affecting on the changes of the classification results. Sub samples were randomly generated based on the total sample size, respectively 1%, 2%, 5%, 10% and 25%.

Then classification results were generated for the five sub samples and analyzed by comparing discriminant analysis and logistic regression models separately. Table 7.5 illustrates the comparison of correct classification results of two models for assigned sub samples.

Table 7.5: Classification Results for Different Sample Sizes

Sample Size (N%)	Percentage of Correct Classification	
	Discriminant Function (%)	**Logistic Regression (%)**
1	78.2	88.2
2	87.6	93.8
3	83.9	94.1
4	86.7	93.7
5	84.7	94.2

It is depicted in table 7.5 that the percentages of correct classification of status of mortality were higher with having higher sub sample sizes, irrespective of the type of model. However, it can be seen that percentage of correct classification is higher for logistic model than the corresponding values in discriminant function. Furthermore, changes of classification accuracy among different sample sizes seems to be negligible for logistic regression model. In contracts, the sample size affects significantly on the classification accuracy for discriminate function. Thus, it can be concluded that logistic model is better than discriminant function.

The overall classification accuracy rate was in the satisfactory level for both modeling techniques. Then the final models can be identified as the best approaches for modeling child mortality.

7.6 Summary of Chapter 7

Of the two models used for modelling the child mortality in Sri Lanka, logistic regression was found as the best fitted model which is having best in predictive ability under the different indicators of used for classification accuracy.

CHAPTER 8

CONCLUSIONS AND SUGGESTIONS

This research is an attempt of addressing the health sector problem by analyzing the data from the demographic health survey in Sri Lanka, focusing on the year 2016. The study objective was modelling the child mortality, aged between 0-5 with the two traditional statistical approaches, linear discriminant analysis and binary logistic regression. Based on the inferences derived from the statistical analyses the following conclusions and suggestions are given below.

8.1 Conclusions

- From both methods it was found that the significant association between child mortality and predictor variables: child mortality, marital status, mother's literacy, status of antenatal care, delivery type, pregnancy duration and decision-making ability.

- Both models found that no significant association between 'status of child mortality' and status of vitamin/medicine usage and birth order.

- The linear discriminant analysis was able to correctly classify 8990 individuals of children out of 10,589 as against the logistic regression was able to classify 9966. The probability of correctly classification by the binary logistic model was 94.1% as against 84.9% by discriminant function.

- In the discriminant analysis, the AER was 15.1% & the ACCR was 84.9% and the logistic regression approach, AER was 5.9% & the ACCR was 94.1%. The ACCR and AER indicating that the model performs well on classification using both statistical approaches. It is concluded that the classification error rates were relatively very close in both methods.

- When it was considered the sensitivity, specificity and the area under curve, difference between logistic regression and discriminant analysis was negligible.

Five sample sizes were selected from the original dataset and performed the similar analysis. It was concluded that variations in sample sizes were not significantly influenced more on classification accuracy and the model performance.

- It is recommended to use binary logistic model to classify the status of mortality of the children of age between 0-5 years. The probability of overall correct classification is 94.1% & compared to 84.9% for linear discriminant function.

8.2 Suggestions

- Modern methods of data modelling are efficient and effective rather than traditional methods. This study results the utilization of two traditional methods and it is encouraged to develop comparative statistical models with modern and traditional methods parallel.

- Boosting health education from the childhood is must in the community development. Therefore, policy makers should responsible in paying attention on female children from the school.

- Establishing community-based healthcare programs and conducting proper awareness programs about the antenatal care, child care, reproduction, personal hygiene etc. are crucially influenced on increasing the child survival and reducing maternal deaths. Educated mothers have a greater ability in protecting their children.

- Poverty and child mortality have a significantly negative relationship in most of the low-income countries. Government should be responsible in facilitating to strengthen the source of income for the poor community and increase the wealth index of the identified society levels.

- Reducing the resource disparity among regions, when allocating health resources in terms of trained staff, health equipment. Reactions to medical causes may be a reason for another death of child or mother. Thus, mothers must be encouraged for the safe delivery at a health facility rather than home delivery.

- Development of on-time accurate information system for collection of health data related to the mother and child care is crucial in reducing child deaths in community level whilst the higher cost in initial investment. When estimating data for policy implications, reliable quality of data improves the effectiveness of the decision-making process.

REFERENCES

Abdi, H. (2007). Discriminant correspondence analysis, In: N.J. Salkind (Ed.): *Encyclopedia of Measurement and Statistic*. Thousand Oaks (CA): Sage. pp. 270–275

Abdulqader, Q.M. (2015). Comparison of discriminant analysis and logistic regression analysis: An application on caesarean births and natural births data, *Journal of The Institute of Natural & Applied Sciences*, Vol.20(1-2), pp.34-46

Adedokun, S.T. Uthman, O.A. Adekanmbi, V.T. and Wiysonge, C.S. (2017). Incomplete childhood immunization in Nigeria: A multilevel analysis of individual and contextual factors, *BMC Public Health*, Vol.17(236), pp.1-10, DOI: 10.1186/s12889-017-4137-7

Adegbosin, A.E. Stantic, B. and Sun, J. (2020). Efficacy of deep learning methods for predicting under-five mortality in 34 low-income and middle-income countries. *BMJ Open*, Vol.10, pp.1-8, DOI:10.1136/ bmjopen-2019-034524

Adetoro, G.W. and Amoo, E.O. (2014). A statistical analysis of child mortality: Evidence from Nigeria, J*ournal of Demography and Social Statistics*, Maiden Edition, pp.110-120

Aheto, J.M.K. (2019). Predictive model and determinants of under-five child mortality: Evidence from the 2014 Ghana demographic and health survey, *BMC Public Health*, Vol.19(64), pp.1-10, [online] Available at: https://doi.org/10.1186/s12889-019-6390-4

Ahmed, L.A. (2017). Using logistic regression in determining the effective variables in traffic accidents, *Applied Mathematical Sciences*, Vol.11(42), pp.2047-2058

Ahmed, S. Koenig, M.A. and Stephenson, R. (2006). Effects of domestic violence on perinatal and early-childhood mortality: Evidence from North India, *American Journal of Public Health*, Vol.96(8), pp.1423-1428

Ahinakwa, L.T. and Oduro, F.T. (2018). Logistic regression model on infant mortality at the Komfo Anokye teaching hospital, Kumasi – Ghana, *International Journal of Probability and Statistics,* Vol.7(3), pp.63-66, DOI: 10.5923/j.ijps.20180703.01

Al-Jawaldeh, A. Abul-Fadl, A. and Tawfik, A. (2018). In-depth analysis of mortality in relation to malnutrition in children under-five of age in the eastern Mediterranean region, *Journal of Nutrition & Weight Loss,* Vol.3(2), pp.1-8

Alayande, S.A. and Bashiru, K.A. (2015). An overview and application of discriminant analysis in data analysis, *IOSR Journal of Mathematics*, Vol.11 (1), pp.12-15, DOI: 10.9790/5728-11151215

Ali, Y.A.E.E. (2005). Child well-being and survival in urban development context, *The case of Khartoum State, International Population Conference*

Antonogeorgos, G. Panagiotakos, D.B. Priftis, K.N. and Tzonou, A. (2009). Logistic regression and linear discriminant analyses in evaluating factors associated with asthma prevalence among 10- to 12-years-old children: divergence and similarity of the two statistical methods, *Int J Pediatr*, Vol.2009, pp.1-6, DOI:10.1155/2009/952042

Attia, S. Versloot, S.J. Voskuijl, W. et al. (2016). Mortality in children with complicated severe acute malnutrition is related to intestinal and systemic inflammation: an observational cohort study, *Inflammation and Mortality in Severe Malnutrition*, Vol.104, pp.1441-1449, DOI: 10.3945/ajcn.116.130518.

Balogun, O.S. Akingbade, T.J. and Oguntunde, P.E. (2015). An assessment of the performance of discriminant analysis and the logistic regression methods in classification of mode of delivery of an expectant mother, *Mathematical Theory and Modeling,* Vol.5(5), pp.147-154

Barclay, K.J. and Kolk, M. (2013). Birth order and mortality: A Population-Based Cohort Study, *IUSSP Conference 2013*

Basu, A.M. and Basu, K. (1991). Women's economic roles and child survival: The case of India, *Health Transition Review,* Vol.1(1), pp.1-20

Becher, H. Muller, O. Jahn, A. Gbangou, A. Wolf, G.K. and Kouyate, B. (2004). Risk factors of infant and child mortality in rural Burkina Faso, *Bulletin of the World Health Organization*, Vol.82, pp.265-273

Benn, C.S. (2012). Combining vitamin, a and vaccines: convenience or conflict? *Danish Medical Journal*, Vol.59(1), pp.1-28

Betz, N.E. (1987). Use of discriminant analysis in counseling psychology research, journal of counseling psychology, Vol.34(1), pp.393-403

Bhaumik, S.K., Discriminant analysis, *Department of Economic Studies & Policy Central University of South Bihar*

Bhatia, M. Dwivedi, L.K. Ranjan, M. et al. (2019). Trends, patterns and predictive factors of infant and child mortality in well performing and underperforming states of India: A secondary analysis using National Family Health Surveys, *BMJ Open*, Vol.9, pp.1-11, DOI:10.1136/bmjopen-2018-023875

Boco, A.G. (2014). Assessing sex differentials in under-five mortality in sub-Saharan Africa: A cross-national comparative analysis, *Canadian Studies in Population*, Prentice Institute, Vol.41(3-4), pp.49-87

Burstein, R. Henry, N.J. et al. (2019). Mapping 123 million neonatal, infant and child deaths between 2000 and 2017, Nature, Vol.574, pp.353-358, [online] Available at: https://doi.org/10.1038/s41586-019-1545-0

Claeson, M. Bos, E.R. Mawji, T. and Pathmanathan, I. (2000). Reducing child mortality in India in the new millennium, *Bulletin of the World Health Organization*, Vol.78(10), pp.1192-1199

Das, A.C. (2015). Childhood mortality and child nutritional status of Bangladesh: A review on demographic and health survey, *Journal of Current and Advance Medical Research*, Vol.2(2), pp.42-46, DOI: 10.3329/jcamr.v2i2.24878

DCS (2014). Department of census and statistics, Sri Lanka, [online] Available at: http://www.statistics.gov.lk/ [Assessed by 25th May 2020]

DHS (2016). Survey Data: Demographic and health survey, *Department of Census and Statistics,* Sri Lanka

Dendup, T. Zhao, Y. and Dema, D. (2018). Factors associated with under-five mortality in Bhutan: An analysis of the Bhutan national health survey 2012, *BMC Public Health*, Vol. 18(1375), pp.1-15, [online] Available at: https://doi.org/10.1186/s12889-018-6308-6

Dissanayake, L. (2000). The Influence of child mortality and breastfeeding on inter-birth intervals in Sri Lanka, *The journal of Family Welfare*, Vol.46(1), pp.21-30

Elgohari, H. (2017). Efficiency of discriminant analysis and multivariate logistic regression for the detection of anemic children with chronic kidney disease, *International Journal of Statistics and Applications*, Vol.7(2), pp.131-136, DOI: 10.5923/j.statistics.20170702.09

El-habil, A. and El-Jazzar, M. (2014). A Comparative Study between linear discriminant analysis and multinomial logistic regression, *An - Najah Univ. J. Res. (Humanities)*, Vol.28(6), pp.1525-1548

Ewbank, D.C. (1994). Maternal education and theories of health behavior: A Cautionary Note, *Health Transition Review*, Vol.4, pp.215-224

Fenta, H.M. Workie, D.L. Zike, D.T. (2020). Determinants of stunting among under-five years' children in Ethiopia from the 2016 Ethiopia demographic and health survey: Application of ordinal logistic regression model using complex sampling designs, *Clinical Epidemiology and Global Health*, Vol.8, pp.404-413, [online] Available at: https://doi.org/10.1016/j.cegh.2019.09.011

F.H.B. (2013). Annual report, family health bureau Sri Lanka, [online] Available at: https://fhb.health.gov.lk/index.php/en/

F.H.B. (2016). Annual report, family health bureau Sri Lanka, [online] Available at: https://fhb.health.gov.lk/index.php/en/

Friendly, M. and Sigal, M. (2018). Visualizing tests for equality of covariance matrices, *Journal of the American Statistician*, Vol.74(2), pp.144-155

Gamhewage, N.C. Weerasekara, M. and Wijayaratne, J. (2018). Pre-discharge neonatal morbidity and mortality in extremely preterm babies managed at Sri Jayewardenepura teaching hospital, *Sri Lanka Journal of Child Health*, Vol.47, pp.155-158

Garenne, M. and Hohmann-Garenne, S. (2003). A wealth index to screen high-risk fantilies: Application to morocco, *J Health Popul Nutr*, Vol.21(3), pp.235-242

Garfield, R. and Leu, C. (2000). A multivariate method for estimating mortality rates among children under 5 years' health & social indicators in Iraq, *International journal of Epidemiology*, Vol.29, pp.510-515

Geneti, K.T. and Deressa, T.K. (2014). Correlates of under-five child mortality in Ethiopia: A logistic regression analysis, *Advances and Applications in Statistics*, Vol.39(1), pp.61-79, [online] Available at: http://pphmj.com/journals/adas.htm

Ghimire, P.R. Agho, K.E. Ezeh, O.K. Renzaho, A.M.N. Dibley, M. and Greenow, C.R. (2019). Under-five mortality and associated factors: Evidence from the Nepal demographic and health survey (2001–2016), International journal of environmental research and public health, Vol.16(1241), pp.1-14, DOI:10.3390/ijerph16071241

Greenberg, B.G. and Wells, H.B. (1963). Linear discriminant analysis in perinatal mortality, Vol.53(4), pp.594-602

Griffis, H.M. (2012). Women's empowerment and infant and child mortality: Incorporating social institutions and context, *Electronic Theses, Treatises and Dissertations*, College of social sciences and public policy, The Florida State university

Gunasekara, D.P.S. and Gunasekara, P.C. (2000). Improving perinatal and neonatal mortality in Sri Lanka: Is it cost effective, *Sri Lanka Journal of Child Health*, Vol.29, pp.23-26

Gurny, P. and Gurny, M. (2013). Comparison of credit scoring models on probability of default estimation for US banks, *Prague Economic Papers,* Vol.2, pp.163-181, DOI: 10.18267/j.pep.446

Hanley, J.A. (1983). Appropriate uses of multivariate analysis, *Ann. Rev. Public Health*, Vol.4, pp.155-180

Hanmer, L. Lensink, R. and White, H., Infant and child mortality in developing countries: Analyzing the data for robust determinants

Hedefalk, F. Quaranta, L. and Bengtsson, T. (2017). Unequal lands: Soil type, nutrition, and child mortality in southern Sweden, 1850–1914, *Demographic Research*, Vol.36(36), pp.1039-1080

Hill, K. (1992). Approaches to the measurement of childhood mortality: A Comparative Review, *Population Index*, Vol.57(3), pp.368-382

Hill, K. (2003). Frameworks for studying the determinants of child survival, *Bulletin of the World Health Organization*, Vol.81(2), pp.138-139

Hossain, M. and Islam,M. (2008). Socio-economic variables affecting infants and children mortality in Bangladesh, *The Internet Journal of Health*, Vol.9(2), pp.1-7

Iqbal, N. Gkiouleka, A. Milner, A. Montag, D. and Gallo, V. (2018). Girls' hidden penalty: Analysis of gender inequality in child mortality with data from 195 countries, *BMJ Global Health*, Vol.3, pp.1-9, DOI:10.1136/bmjgh-2018-001028

Islam, R. Hossain, M. Rahman, M. and Hossain, M. (2013). Impact of socio-demographic factors on child mortality in Bangladesh: An multivariate approach, *International Journal of Psychology and Behavioral Sciences*, Vol.3(1), pp.34-39, DOI: 10.5923/j.ijpbs.20130301.05

Jayadevan, C.M. (2016). Child mortality rates in ASIA, *Scholars Journal of Applied Medical Sciences*, Vol.4(5A), pp.1491-1505

Kaufmann, G. and Cleland, J. (1994). Maternal education and child survival: anthropological responses to demographic evidence, *Health Transition Review*, Vol.4, pp.196-199

Kembo, J. and Ginneken, J.K.V. (2009). Determinants of infant and child mortality in Zimbabwe: Results of multivariate hazard analysis, *Research Article*, Vol.21(13), pp.367-384, DOI: 10.4054/DemRes.2009.21.13

Khan, J.R. and Awan, N. (2017). A comprehensive analysis on child mortality and its determinants in Bangladesh using frailty models, *Archives of Public Health*, Vol.75(58), DOI 10.1186/s13690-017-0224-6

Koenig, M.A. Khan, M.A. Wojtyniak, B. Clemens, J.D. Chakraborty, J. Fauveau, V. Phillips, J.F. Akbar, J. and Barua, U.S. (1990). Impact of measles vaccination on childhood mortality in rural Bangladesh, *Bulletin of the World Health Organization*, Vol.68(4), pp.441-447

Krzanowski, W. J. (1986). Multiple discriminant analysis in the presence of mixed continuous and categorical data, *Computers and Mathematics with Applications*, Vol.12A(2), pp.179–185.

Liu, K. Moon, M. Sulvetta, M. and Chawla, J. (1992). International infant mortality rankings: A look behind the numbers, *Health care Financial Review*, Vol.13(4), pp.105-118

Lundberg, E. and Svaleryd, H. (2016). birth order and child health, *Labour Market and Discrimination Study*

Mahmoud, O.H. (2008). A Multivariate model for predicting the efficiency of financial performance for property and liability Egyptian insurance companies, *Casualty Actuarial Society, Discussion Paper Program*

Maiprasert, D. and Kitbumrungrat, K. (2012). Comparison multinomial logistic regression and discriminant analysis in predicting the stage of breast cancer, *IJCSNS International Journal of Computer Science and Network 44 Security*, Vol.12(4), pp.44-52

Makepeace, G. and Pal, S. (2006). Understanding the effects of siblings on child mortality: Evidence from India, *Discussion Paper*, No.2390

Maniruzzaman, M.D. Suri, H.S. Kumar, N. (2018). Risk factors of neonatal mortality and child mortality in Bangladesh, Journal of Global Health, Vol.8(1), pp.1-16, DOI: 10.7189/jogh.08.010421

Memic, D. (2015). Assessing credit default using logistic regression and multiple discriminant analysis, *Interdisciplinary Description of Complex Systems*, Vol.13(1), pp.128-153, DOI: 10.7906/indecs.13.1.13

Masuy-Stroobant, G. (2001). The determinants of infant mortality: how far are conceptual frameworks really modelled? *Belgium: Université catholique de Louvain*

Montgomery, M.E. White, M.E. and Martin, S.W. (1987). A Comparison of discriminant analysis and logistic regression for the prediction of coliform mastitis in dairy cows, *Can J Vet Res*, Vol.51, pp.495-498

Mosley, W.H. and Chen, L.C. (1984). An analytic framework for the study of child survival in developing countries, *Population and Development Review*, Vol.10, pp.25-45

Morris, S.S. Black, R.E. and Tomaskovic, L. (2003). Predicting the distribution of under-five deaths by cause in countries without adequate vital registration systems, *International Journal of Epidemiology*, Vol.32, pp.1041-1051, DOI: 10.1093/ije/dyg241

Mmusi-Phetoe, R.M.M. (2016). Social factors determining maternal and neonatal mortality in South Africa: A qualitative study, *AOSIS*, Vol.39(1), pp.2223-6279, [online] Available at: http://dx.doi.org/10.4102/curationis.v39i1.1571

Nobles, J. Frankenberg, E. and Thomas, D. (2015). The effects of mortality on fertility: population dynamics after a natural disaster,

Pham, T.L. Kooreman, P. Koning, R.H. and Wiersma, D. (2011). Gender patterns in Vietnam's child mortality, *IZA Discussion Paper*, No.5741

Pillai, V.K. Maleku, A. and Wei, F.H. (2013). Maternal mortality and female literacy rates in developing countries during 1970–2000: A latent growth curve analysis, *International Journal of Population Research*, [online] Available at: http://dx.doi.org/ 10.1155/2013/163292

Poulsen, J. and French, A., Discriminant function analysis (DA), *DocPlayer*, [online] Available at: http://www.google.com/amp/s/docplayer.net

Rahman, S. Rahman, S. and Rahman, A. (2019). Determinants of death among under-5 children in Bangladesh, *MPRA Paper*, No.93511, [online] Available at: https://mpra.ub.uni-muenchen.de/93511/

Rajindrajith, S. Mettananda, S. Adihetti, D. Goonawardana, R. and Devanarayana, N.M. (2009). Neonatal mortality in Sri Lanka: Timing, causes and distribution, *The Journal of Maternal-Fetal & Neonatal Medicine*, Vol.22(9), pp.791-796

Rawlings, S. and Siddique, Z. (2018). Domestic violence and child mortality, *Discussion Paper Series*, IZA Institute of Lab or Economics,

Rice, A.L. Sacco, L. Hyder,A. and Black, R.E. (2000). Malnutrition as an underlying cause of childhood deaths associated with infectious diseases in developing countries, *Bulletin of the World Health Organization*, Vol.78(10), pp.1207-1221

Samarakoon, S. Madurawela, S. and Bandara, S. (2015). Chapter 9: Health sector reforms, *Sri Lanka State of the Economy Report*, pp.133-150

Sanghera, T. (2018). India's under-5 mortality now matches global average, But Bangladesh, Nepal do better, *IndiaSpend*, [online] Available at: https://www.indiaspend.com

Save the Children (2017). Sri Lanka supplementary report on the United Nations convention on the rights of the child

Senanayake, H. Goonewardene, M. Ranatunga, A. Hattotuwa, R. Amarasekara, S. and Amarasinghe, I. (2011). Achieving millennium development goals 4 and 5 in Sri Lanka, *International Journal of Obstetrics and Gynaecology*, pp.78-87, DOI: 10.1111/j.1471-0528.2011.03115.

Shann, F. (2013). Nonspecific effects of vaccines and the reduction of mortality in children, *Clinical Therapeutics*, Vol.35(2), pp.109-114

Sharifzadeh, G.R. Namakin, K. and Mehrjoofard, H. (2008). An epidemiological study on infant mortality and factors affecting it in rural areas of Birjand, Iran, *Iran J Pediatr*, Vol.18(4), pp.335-342

Shetty, A. and Shetty, S. (2014). The Impact of female literacy on infant mortality rate in Indian States, *Curr Pediatr,* Vol.18(1), pp.49-56

Silva, R.D. (2015). Realising the future we want for all our children, *Sri Lanka Journal of Child Health*, Vol.44(4), pp.186-192

Silverman, J.G. Decker, M.R. Cheng, D.M. Wirth, K. Saggurti, N. McCauley, H.L. Falb, K.L. Donta, B. and Raj, A. (2011). Gender-based disparities in infant and child mortality based on maternal exposure to spousal violence, *Arch Pediatr Adolesc Med.*,Vol.165(1), pp.22-27, DOI:10.1001/archpediatrics.2010.261.

SLDHS (2016). Annual report on Sri Lanka demographic and health survey, [online] Available at: http://www.statistics.gov.lk/ [Assessed by 15th May 2020]

Sufian, A.J.M. (2013). Life expectancy and its socioeconomic determinants: A discriminant analysis of national level data*, International Journal of Humanities and Social Science,* Vol.3(12), pp.303-312

Tagoe, E.T. Agbadi, P. Nakua, E.K. (2020). A predictive model and socioeconomic and demographic determinants of under-five mortality in Sierra Leone, *Heliyon*, Vol.6, pp.1-7, https://doi.org/10.1016/j.heliyon.2020.e03508

Thakrar, A.P. Forrest, A.D. Maltenfort, M.G. and Forrest, C.B. (2018). Child mortality in The US and 19 OECD comparator nations: A 50-year time-trend analysis, *Health Affairs*, Vol.37(1), pp.140-149

Thakur, J. and Pahuja, S.K. (2019). Performance comparison of prediction models for neonatal sepsis using logistic regression, multiple discriminant analysis and artificial neural network, *Biomedical Physics & Engineering Express*, Vol.5(3), DOI: 10.1088/2057-1976/aaf677

Tuti, T. Agweyu, A. Mwaniki, P. (2017). An exploration of mortality risk factors in non-severe pneumonia in children using clinical data from Kenya, *BMC Medicine*, Vol.15(201), pp.1-12, DOI: 10.1186/s12916-017-0963-9

United Nations (2017). Levels & trends in child mortality: Estimates Developed by the United Nation's Inter-Agency Group for Child Mortality Estimation, [online] Available at: https://data.unicef.org [Assessed by 5th June 2020]

UNICEF (2017). Report on child mortality, https://www.unicef.org/publications/index_102899.html, [Assessed by 25th May 2020]

UN IGME (2018). UNICEF Data: Estimates developed by the United Nations inter-agency Group for Child Mortality Estimation, [online] Available at: https://data.unicef.org [Assessed by 25th May 2020]

Vidal-e-Silva, S.M.C. Tuon, R.A. Probst, L.F. Gondinho, B.V.C. Pereira, A.C. Meneghim, M.C. et al. (2018). Factors associated with preventable infant death: a multiple logistic regression, *Rev Saude Publica*, Vol.52(32), pp. 1-11, [online] Available at: https://doi.org/10.11606/S1518-8787.2018052000252

Vine, R.A.L. Dexter, E. Velasco, P. Vine, S.L. Joshi, A.R. Stuebing, K.W. and Tapia-Uribe, F.M. (1994). Maternal literacy and health care in three countries: a preliminary report, *Health Transition Review*, Vol.4, pp.186-191

Weerasekara, M. Amarasekara, S. and Siriwardhane, H.V.D.N. (2019). An analysis of neonatal mortality in a tertiary care hospital over a decade-2008 to 2017, *Sri Lanka Journal of Child Health*, Vol.48(4), pp.326-333

World Bank (2018). Open Data, [online] Available at: https://data.worldbank.org/ [Assessed by 25th May 2020]

Zewdie, S.A. and Adjiwanou, V. (2017). Multilevel analysis of infant mortality and its risk factors in South Africa, *International Journal of Population Studies*, Vol.3(2), pp.43-56

Zylbersztejn, A. Gilbert, R. Hjern, A. Wijlaars, L. and Hardelid, P. (2018). Child mortality in England compared with Sweden: A birth cohort study, *Lancet*, Vol.391, pp.2008-2018