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

Abstract iii

Acknowledgement iv

Declaration of the Candidate v

Declaration of the Supervisor vi

Table of Contents vii

List of Figures x

List of Tables xi

List of Acronyms xiv

1.0. Chapter 1 – Introduction 1-13

1.1. Background of the Study 1

1.2. Child Mortality in Sri Lanka 2

1.3. Child Mortality in Global Context 6

1.4. Life Expectancy at Birth 9

1.5. Importance of Fertility and Mother Care 10

1.6. Research Problem 11

1.7. Objectives of the Study 12

1.8. Significance of the Study 12

1.9. Outline of the Dissertation 13

2.0. Chapter 2 – Literature Review 14-28

2.1. Child Mortality in Sri Lanka 14

2.2. Global Child Mortality 15

2.3. Common Risk Factors of Child Mortality 17

2.4. Social Responsibility of Reducing Child Mortality 22

2.5. Review of Classification Models 23

2.6. Comparison of DA & LR Models 26

2.7. Summary of Chapter 2 28

3.0. Chapter 3 – Materials & Methods 29-45

3.1. Population & Sample 29
3.2. Sampling Procedure 29
3.3. Methods of Data Collection and Analysis 29
3.4. Conceptual Framework 30
3.5. Selection of Variables 31
3.6. Key Measurements 33
3.7. Dummy Coding of Selected Variables 34
3.8. Univariate Test Statistics 34
3.9. Discriminant Analysis 36
3.10. Binary Logistic Regression 39
3.11. Model Evaluation 42

4.0. Chapter 4 – Analysis of Risk Factors 46-53

4.1. Behaviour of Risk Factors 46
4.2. Risk Factors influenced on Childhood Mortality 48
4.3. Summary of Chapter 4 53

5.0. Chapter 5 – Development of Linear Discriminant Model 54-65

5.1. Development of Linear Discriminant Model 54
5.2. Estimation of Canonical Discriminant Function 58
5.3. Estimation of Fisher's Linear Discriminant Function 61
5.4. Validation of the Developed Model 63
5.5. Predictive Power of the Model 64
5.6. Summary of Chapter 5 64

6.0. Chapter 6 – Development of Binary Logistic Regression Model 66-74

6.1. Binary Logistic Model with all Eight Variables 66
6.2. Development of Model via Stepwise Forward (LR-Ward/Likelihood) 69
6.3. Development of Model via Stepwise Backward(LR-Likelihood) 72

6.4. Predicted Probability of Logistic Function 73
6.5. Predictive Power of the Model 74
6.6. Summary of Chapter 6 74

7.0. Chapter 7 – Comparison of Two Models 75-82

7.1. Identification of Significant Variables 75
7.2. Probability of Misclassification 76
7.3. Receiver Operating Characteristic (ROC) Curve 77
7.4. Analysis of Sensitivity/ Specificity 80
7.5. Comparison of Accuracy with Sample Size Variations 81
7.6. Summary of Chapter 7 82

8.0. Chapter 8 – Conclusions and Suggestions 83-85

8.1. Conclusions 83
8.2. Suggestions 84

References..... 86-96

LIST OF FIGURES

Chapter 1 – Introduction

Figure 1.1: District Disparities in Live Births & Infant Deaths, 1997-2014 5

Chapter 3 – Materials & Methods

Figure 3.1: Conceptual Framework of the Research 31

Chapter 5 – Development of Linear Discriminant Model

Figure 5.1: Distribution of the Canonical Discriminant Scores 60

Figure 5.2: Distribution of the Fisher's Discriminant Scores for LDF1 62

Figure 5.3: Distribution of the Fisher's Discriminant Scores for LDF2 63

Chapter 6 – Development of Binary Logistic Regression Model

Figure 6.1: Distribution of Predicted Probability among Two Groups 73

Chapter 7 – Comparison of Models

Figure 7.1: ROC Curve for Discriminant Function 78

Figure 7.2: ROC Curve for Logistic Regression Model 79

Figure 7.3: Comparison of ROC Curves for Sample Size, n = 5% of Total 80

LIST OF TABLES

Chapter 1 – Introduction

Table 1.1: Age Specific Childhood Mortality Rates in SL, 2013-2018..... 2

Table 1.2: Under-Five Mortality Rate in SL, by Gender,2016..... 3

Table 1.3: Under-Five Mortality Rate in SL, by BO & ME,2016..... 4

Table 1.4: Under-Five Mortality Rates by Selected SDG Regions, 2012-2018..... 7

Table 1.5: Globl Distribution of Child Deaths by Causes, 2016..... 8

Table 1.6: Life Expectancy at Birth by Highlighted Countries, 2012-2018..... 9

Table 1.7: Fertility Rates vs Infant Mortality Rate, 2012-2018..... 10

Chapter 3 – Materials & Methods

Table 3.1: Detailed Description of the Selected Variables..... 32-33

Table 3.2: Contingency Table..... 43

Table 3.3: Representation of Error Classification..... 43

Table 3.4: Representation of Sensitivity/Specificity..... 44

Chapter 4 – Analysis of Risk Factors

Table 4.1: Frequency Distribution of Response Variable..... 47

Table 4.2: Frequency Distribution of Categorical Variables..... 47

Table 4.3: Association between the Status of Mortality and Gender..... 48

Table 4.4: Association between the Status of Mortality and Birth Order..... 48

Table 4.5: Association between the Status of Mortality and Marital Status..... 49

Table 4.6: Association between the Status of Mortality and Mother's Literacy 50

Table 4.7: Association between the Status of Mortality and Status of Antenatal Care..... 50

Table 4.8: Association between the Status of Mortality and Status of Vitamin/ Medicine Usage 51
Table 4.9: Association between the Status of Mortality and Delivery Type 51
Table 4.10: Association between the Status of Mortality and Pregnancy Duration .. 52
Table 4.11: Association between the Status of Mortality and Decision Making Ability 52

Chapter 5 – Development of Linear Discriminant Model

Table 5.1: Significance of Discriminating Variables 55
Table 5.2: Results of Stepwise Method in DA 56-57
Table 5.3: Variables not in the Analysis based on the DA 57-58
Table 5.4: Unstandardized Coefficients of Canonical Discriminant Analysis 58
Table 5.5: Correlation Matrix of the Independent Variables 59
Table 5.6: Coefficients of Canonical Function 61
Table 5.7: Coefficients of FLD in Stepwise DA 61
Table 5.8: Significance Test of Predictors in Canonical Function 64
Table 5.9: Classification Results of the Discriminant Function 64

Chapter 6 – Development of Binary Logistic Regression Model

Table 6.1: Omnibus Test of Model Coeffients 66
Table 6.2: Model Summary 67
Table 6.3: Hosmer & Lemeshow Test Statistics 67
Table 6.4: Properties of the Coefficients of the fitted Model 68
Table 6.5: Results of Hosmer and Lemeshow Test under Stepwise Forward(Wald/Likelihood) 69

Table 6.6: Variables in the Model Obtained in Stepwise Forward (Wald/Likelihood) 69-71

Table 6.7: Results of Hosmer and Lemeshow Test under Stepwise Backward(Likelihood) 72

Table 6.8: Properties of the Coefficients of the Model developed via Backward Likelihood 72-73

Table 6.9: Classification Results of Stepwise Logistic Regression Model..... 74

Chapter 7 – Comparison of Two Models

Table 7.1: Summary Steps in Predictors Added/ Removed..... 75

Table 7.2: Performance Summary with Misclassification 76

Table 7.3: Test Results of Area Under Curve 77

Table 7.4: Analysis of Sensitivity and Specificity 81

Table 7.5: Classification Results for Different Sample Sizes 81

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