Factors Affecting Child Mortality in Lesotho using 2009 and 2014 LDHS Data

By

Nonduduzo Noxolo Mkhize

A thesis submitted in fulfilment of the requirement of the degree for the

MASTER OF SCIENCE

in

STATISTICS

Under the supervision of

Dr S Melesse,

And Co-Supervision of

Prof H Mwambi

And

Prof S Ramroop

UNIVERSITY OF KWAZULU-NATAL
SCHOOL OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE
PIETERMARITZBURG CAMPUS, SOUTH AFRICA.

Declaration

I, Nonduduzo Noxolo Mkhize, declare that this thesis titled, 'Factors Affecting Child Mortality in Lesotho(2009,2014 LDHS DATA) and the work presented in it is my own. I confirm that:

- This work was done while in candidature for a master's degree at the University.
- The thesis has not been submitted for any degree or qualification at any other university.
- I have duly acknowledged the published work of other researchers and referenced the sources of pictures, graphs, and data used in the research work.
- Where I have quoted from the work of others, the source is always given.

I hereby confirm that all passages which are literally or in general matter taken out of publications or other sources are marked as such.

The research work described in this thesis was carried out in the School of Mathematics, Statistics and Computer Science , University of KwaZulu-Natal, Pietermaritzburg Campus, under the supervision of Dr. Sileshi Melesse and the co-supervision of Prof H. Mwambi and Prof S.Ramroop. The work represents original work by the author and has not otherwise been submitted in any form for any degree or diploma to any University. Where use of the work of others has been made it is duly acknowledged.

[redacted]	16/09/2021
Miss. Nonduduzo N. Mkhize	Date
[redacted]	15/09/2021
Prof. Sileshi F. Melesse	Date
[redacted]	15/09/2021
Prof. H. Mwambi	Date
[redacted]	15/09/21
Prof. S. Ramroop	Date

Acknowledgements

Firstly, I would like to say thank you to everyone who supported me throughout this whole process whether I have mentioned or not. Thank God for granting me the power and discipline to finish the research successfully.

Thank you to my supervisor Dr. Sileshi Melesse for his guidance, patience, encouragement, and supervision during the period of my study. Thank you for believing in me and inspiring me that I can do this. I would also like to thank my co-supervisor Prof H Mwambi and Prof.Ramroop for their contributions. My special thanks to the Demographic and Health Survey (DHS) (Lesotho) for allowing me to use the data utilized in analysis for this thesis.

I thank the staff members in the School of Mathematics , Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, for the study environment they have provided. Everyone who contributed greatly to the success of this thesis either directly or indirectly though not mentioned by name deserves my thanks. To my best friend Mlungisi for being the best. I could always count on you for support, courage, faith, love, cheers and prayer. To God almighty, thank you for giving me strength and faith.

Finally, I would like to extend my deepest gratitude to my family without their love, support and understanding I could never have completed this degree.

Abstract

Child mortality rate is known to be the important indicator of social development, quality of life, welfare as well as the overall health of the society. In most countries, especially the developing countries; the death of a child is usually caused by transferable, preventable diseases and poor health. Progress in improving under-five mortality since 1990 has been made globally. There has been a decline globally in under-five mortality from 12.7 million in 1990 to approximately 6 million in 2015. All regions except the developing countries in Sub-Saharan Africa, Central Asia, Southern Asia and Oceania had reduced the rate by 52% or more in 2013. Lesotho is a developing country with one of the highest rates of infant and child mortality.

The study uncovers the factors influencing child mortality in Lesotho based on the Lesotho Demographic and Health Surveys for 2009 and 2014. The survey logistic regression, a model under the generalized linear model framework was used to find the factors related to under-five child mortality to account for the sampling designs complexity. The SLR model is not able to account for variability occurring from connection between subjects from the equal clusters and household. The generalized linear mixed model is then put into application. To ease the normality assumptions and the linearity assumption in the parametric models, the semi-parametric generalized additive model, was lastly used for the data. Finding the determining factors that result in child mortality will benefit the way intervention programs are planned and the formulation for policy makers to lead in the decreasing of child mortality; and accomplish MDGs. This study intends to improve the existing knowledge on child mortality in Lesotho by studying the determining factors in detail. Based on the previous studies this paper will recommend intervention designs and policy formulation. Overall, the findings of this research showed that birth order number, weight of child at birth, age of child, breastfeeding, wealth index , education attainment, mother's age, type of place of residence, number of children living were the key determining factors of the under-five mortality in Lesotho. The study displays that policy makers should strengthen the interventions for child health in order to decrease child under-five mortality. The results achieved can help with the policy formulation to control and reduce child mortality. The government should continually assess current programs to review and develop programs that are more applicable.

Table of Contents

Declaration .... **ii**
Acknowledgements .... **iii**
Abstract .... **iv**
Abbreviations .... **xi**
1 Introduction **1**
1.1 Background 1
1.2 Country Background: Lesotho 3
1.3 Health issues, Health Budget and Poverty Profile 6
1.4 The importance of the study 7
1.5 Objectives 8
1.6 Data and Methods 8
1.6.1 Description of the Data 8
1.6.2 Methodology 9
1.7 Outline of this study 9
2 Literature Review **11**
3 Exploratory Data Analysis **16**
3.1 Introduction 16
3.2 Study Variables 16
3.2.1 Response Variable 16
3.2.2 Exploratory Variables 16
3.3 Preliminary Analysis 17
3.4 Summary 23
4 Generalized Linear Models **24**
4.1 Introduction 24
4.2 The Generalized Linear Model 24
4.2.1 Exponential Family of Distributions 25
4.2.2 Components of the GLMs 26
4.2.3 Maximum Likelihood Estimation (MLE) 26
4.3 Model Selection and Diagnostics 27
4.3.1 Model Selection 27
4.3.2 Model Checking 29

4.4 Logistic regression 30
4.4.1 Logistic Regression Model 31
4.4.2 Assumptions of Logistic Regression 32
4.4.3 Limitations to the Logistic Regression 33
4.4.4 Parameter Estimation (MLE) 34
4.4.5 Newton-Raphson Method 36
4.4.6 Parameter Estimation 37
4.5 Logistic Model Selection and Checking 40
4.5.1 Model Selection 40
4.5.2 Model Checking 42
4.6 Fitting the Logistic Regression Model 46
4.6.1 Fitting a LR model with a dichotomous response 46
4.6.2 Multiple Logistic Regression Model 50
4.6.3 Prediction Accuracy of the Logit Model 56
4.6.4 Logistic Regression Diagnostic Plots 58
4.7 Survey Logistic Regression Model 59
4.7.1 Estimation of Parameters 60
4.7.2 Survey Logistic Model Selection and Checking 62
4.7.3 Design Effects 63
4.8 Results of the Survey Logistic Regression Model 65
4.8.1 Model Checking and Prediction Accuracy of the Model 65
4.8.2 Interpretation of the Coefficients of the Model and Odds ratio 68
4.8.3 Comparing the Logistic and Survey Logistic Regression 75
4.9 Shortcomings of the SURVEYLOGISTIC Procedure 79
5 Generalized Linear Mixed Models 81
5.1 Introduction 81
5.2 Generalized Linear Mixed Model 81
5.3 Model Formulation 83
5.4 Parameter Estimation 84
5.4.1 Penalized Quasi-Likelihood 85
5.4.2 Marginal Quasi-Likelihood 87
5.4.3 Laplace Approximation 88
5.4.4 Gaussian Quadrature 89

5.5 Generalized Linear Mixed Models (GLIMMIX) 90
5.5.1 Application of Generalized Linear Mixed Model to the data 90
5.5.2 Summary of Generalized Linear Mixed Models 102
6 Semi Parametric Regression Approach 103
6.1 Introduction 103
6.2 Generalized Additive Models 104
6.2.1 Additive Models 105
6.3 Smoothing 105
6.3.1 What is a Smoother? 105
6.3.2 Cubic Smoothing Splines 106
6.4 Choosing Smoothing Parameters λ 108
6.4.1 Average Mean Square and Predictive Square Error 108
6.4.2 Cross Validation(CV) 109
6.4.3 Generalized Cross Validation(GCV) 109
6.4.4 Degrees of Freedom of a Smoother 110
6.5 Fitting the Generalized Additive Models 110
6.5.1 Back-fitting Algorithm 111
6.5.2 The General Local Scoring Algorithm 112
6.6 Spline Bases and Penalties 114
6.7 P-IRLS (Penalized Iteratively Re-Weighted Least Squares) 116
6.8 Generalized Additive Logistic Regression Model and Application 118
6.9 Fitting the GAM Model using the GAM PROC 118
6.9.1 Fitting the Model 119
7 Discussion 124
7.1 Conclusions and Recommendations 124
7.2 Limitations 127
Appendix A 128
Codes Used to fit the models 128
A.1 SAS code for 2009 LDHS: Logistic Regression 128
A.2 SAS code for 2009 LDHS: Survey Logistic Regression 128
A.3 SAS code for 2009 LDHS: Generalized Linear Mixed Model 129
A.4 SAS code for 2009 LDHS: Generalized Additive Model 129
A.5 SAS code of 2014 LDHS: Logistic Regression 129

A.6 SAS code of 2014 LDHS: Survey Logistic Regression 130
A.7 SAS code of 2014 LDHS: Generalized Linear Mixed Model 130
A.8 SAS code of 2014 LDHS: Generalized Additive Model 130
References 131

List of Tables

Table 3.1: Description of study variables. 16
Table 3.2: Summary of selected characteristics of children (LDHS 2009). 18
Table 3.3: Summary of selected characteristics of children (LDHS 2014). 20
Table 4.1: Classification table.......... 43
Table 4.2: Logistic regression model fit statistics for binary response for LDHS 2009 data. 46
Table 4.3: Logistic regression model fit statistics for binary response for LDHS 2014 data. 46
Table 4.4: Model evaluation for the LR Model LDHS 2009 data. 46
Table 4.5: Model evaluation for the LR Model LDHS 2014 data. 48
Table 4.6: Type 3 Analysis of Effects for the Logistic Regression for LDHS 2009 data.......... 48
Table 4.7:Type 3 Analysis of Effects for the Logistic Regression for LDHS 2014 data.......... 49
Table 4.8: Logistic regression model coefficients, standard errors, odds ratios, confidence intervals and significant two-way interaction effects for LDH 2009 data.......... 50
Table 4.9: Logistic regression model coefficients, standard errors, odds ratios, confidence intervals and significant two-way interaction effects for LDHS 2014 data. 53
Table 4.10: Statistics for Model fit for the survey logistic regression(LDHS 2009). 65
Table 4.11: Statistics for Model fit for the survey logistic regression (LDHS 2014). 66
Table 4.12: Model evaluation for Survey logistic regression (LDHS 2009). 66
Table 4.13: Model evaluation for Survey logistic regression (LDHS 2014). 66
Table 4.14: Type 3 analysis of effects for the LR with complex survey design (2009). 67
Table 4.15: Type 3 analysis of effects for the LR with complex survey design (2014). 68
Table 4.16: Survey logistic regression model estimation for main effects and significant interaction effects(2009). 68
Table 4.17: Survey logistic regression model estimation for main effects and significant interaction effects(2014). 71
Table 4.18: Design effects estimated for response variable (2009). 75
Table 4.19:Design effects estimated for response variable (2014). 77
Table 5.1:Type 3 analysis using 2009 data. .... **Error! Bookmark not defined.**
Table 5.2: Type 3 analysis using 2014 data. .... **Error! Bookmark not defined.**
Table 5.3: Estimates and odds ratios (OR) with 95% confidence intervals for the fixed effects. (2009 DHS DATA) 93
Table 5.4: Estimates and odds ratios (OR) with 95% confidence intervals for the fixed effects. (2014 DHS DATA) 96
Table 6.1: Analytical information about the fitted model(2009) 120
Table 6.2: Analytical information about the fitted model(2014) 120
Table 6.3: Analysis of Deviance 121
Table 6.4: Analysis of Deviance 122

List of Figures

Figure 1.1: Map of Lesotho 4
Figure 1.2:Most populated cities in Lesotho........ 4
Figure 2.1: Groups of proximate determinants 15
Figure 3.1: Sample distribution of children for each region in Lesotho for the two periods. 22
Figure 3.2: Percentage of child survivals for 2009 and 2014 in Lesotho........ 22
Figure 3.3: Wealth Index of Mother in Lesotho (2009 and 2014). 23
Figure 4.1: Sensitivity against False positive rate........ 45
Figure 4.2: (ROC) Receiver Operating Characteristic curve for logit model. 57
Figure 4.3: (ROC)Receiver Operating Characteristic curve for logit model. 57
Figure 4.4: Influence diagnostic from the logistic regression model (2009). 58
Figure 4.5: Influence diagnostic from the logistic regression model (2014) 59
Figure 5.1: Diffogram for breastfeeding by current age of child.(2009 DHS DATA). 99
Figure 5.2: Diffogram for breastfeeding by current age of child.(2014 DHS DATA). 99
Figure 5.3: Analysis of means for breastfeeding by current age of child interaction effects(2009). 100
Figure 5.4: Analysis of means for breastfeeding by current age of child interaction effects(2014). 100
Figure 5.5: Diffogram for Region effect(2009). 101
Figure 5.6: Diffogram for Region effect(2014). 101
Figure 5.7: Analysis of means for Region effect(2009). 102
Figure 5.8: Analysis of means for Region effect(2014). 102
Figure 6.1 107
Figure 6.2 107
Figure 6.3 107
Figure 6.4: Partial Prediction for Each Predictor (2009). 122
Figure 6.5: Partial Prediction for Each Predictor (2014). 123

Abbreviations

AIC - Akaike Information Criterion

AIDS - Acquired Immuno-Deficiency Syndrome

CI - Confidence Interval

GAM – Generalized Additive Models

GLM - Generalized Linear Models

GLMM - Generalized Linear Mixed Models

HIV -Human Immunodeficiency Virus

LDHS – Lesotho Demographic and Health Survey

LR – Logistic Regression

MDGs- Millennium Development Goals

MLE - Maximum Likelihood Estimate

OR - Odds Ratio

ROC -Receiver Operating Characteristic Curve

SAS - Statistical Analysis Software

SLR-Survey Logistic Regression

SSA- Sub Saharan Africa

WHO - World Health Organization

MOHSW - Ministry of Health and Social Welfare

BOS - Bureau of Statistics

Chapter 1

1 Introduction

1.1 Background

It is crucial for evaluating the child well-being, health status as well as the social well-being of any country to assess child mortality rate. In many countries, the top indicator of the overall development and level of child health is the child mortality rate (Rutstein, 2000). It can also be an indicator for coverage of economic development and social development as well as child survival interventions. This makes it important to determine factors that affect child mortality for more development in countries.

Everyday approximately 15,000 children below the age of five die out of 375000 which is equal to 4% globally. Former research illustrates that six million over eleven million children who die every year can be saved by effective measures like giving micronutrient supplements, enhanced family care, vaccines, insecticide-treated bed nets, antibiotics and breastfeeding practices (Animaw et al., 2014).

Since 1990, the world has seen considerable progress in decreasing the death of children. Nevertheless, in order to obtain additional progress, there should be more efforts to assist in the reduction of child mortality. The child mortality rate reduced by over 50 per cent from 1990 to 2015 globally, with significant acceleration in reduction (You et al., 2015)

Infant mortality has been declining worldwide, yet millions of infants die each year (Ahmad et al., 2000). More than 90% of these infant dying occur in the developing countries of the world (Leowski, 1986). Amongst these deaths, sub-Saharan holds the highest rates while South-central Asia contains the highest number of neonatal deaths. Majority of the deaths that happen in these regions are usually preventable (Tan, 2014).

Countries in the Sub Saharan Africa have the highest infant mortality rates in the world, having infant mortality equal to 104 deaths per 1000 live births in 1990's, whilst the rate was 71 deaths per 1000 live birth for the rest of less developed countries (Kalipeni, 2000). There has been a reduced child mortality rate in the Sub-Saharan African since the 1990's and most of these improvements were achieved by interventions that targeted the communicable diseases (Claeson & Waldman, 2000). Health transition beginning later compared to other parts of the world in the sub Saharan Africa(SSA) is one of the factors that explains why child mortality is high in SSA (Garenne & Gakusi, 2006).

Findings on a study by (Omariba et al., 2008) signifies that socio economic factors and socio cultural factors highly influenced the high infant mortality in the sub Saharan Africa. Although a vast literature is of existence on child mortality, evidence on the rates stay elevated in the Sub-Saharan African countries in spite of action strategies and intervention programs done remains insufficient. Since 2000, the taking on of the MDGs, which aimed to decrease the level of child mortality by two-thirds of their levels in 1990 by 2015, the rate of child mortality has reduced globally (Gaffey et al., 2015). One of the things that contributed more to the decrease in child mortality are awareness of the mother ,the interventions of disease-oriented programs, environmental factors, and socioeconomic factors of the household. These are known to be crucial factors towards reducing child mortality rate. The child mortality rate has dropped from 90 per live births to 43 deaths per 1,000 live births between the year 1990 and 2015. Regardless the growing population in the developing countries, the number of children dying has declined from 12.7 million in 1990 to approximately 5.9 million in year 2015. With this progression in the last two decades in child mortality, improvement was not sufficient to meet the MDGs (Organization et al., 2015), "Millennium Development Goal #4 is to reduce child mortality rate by two-thirds".

The study aims to analyze under-five mortality in Lesotho using existing statistical methods, identify what affects it, and what can be done in future to decline the rate of under-five mortality. "Under-five mortality is defined as the probability of a child born in a specific year or period dying before reaching the age of five." (Ahmad et al., 2000). This study identifies the relationship and the effects of some factors like the wealth index, the mother's education, region, sex of the child, mother's age, place of residence, as well as breastfeeding on child mortality in Lesotho for years

2009 and 2014. It explains the rate of child mortality through logistic regression analysis as a function of different factors affecting child mortality. It will also look at generalized linear mixed models and generalized additive model analysis.

Logistic regression analysis is put in use to find the significance of each factor on child mortality. This information can later be used to propose areas for policy makers to take into consideration in addressing child mortality to bring more levels of development.

1.2 Country Background: Lesotho

Lesotho is a country found in the Southern portion of Africa. It is bounded by the Republic of South Africa(RSA), on which it must depend on for access to the outside world (Clarke & Casey, 1995). This country is a small landlocked country with many mountains and borders on the provinces- Free State, KwaZulu-Natal, and Eastern Cape (Steinberg, 2005).

The estimated population in 2019 of Lesotho is 2.2 million, ranking number 145 in the whole world. This in comparison to the 2004 population of 1.9 million and 2009 population of 2.0 million (Population, 2019). The country of Lesotho has 10 administrative regions with total area of 30355 square kilometers. Since the independence in 1966, there has been population movement towards the urban capital city, Maseru, which is the largest city in Lesotho. Smaller urban populations dwell in the regions Maputsoe, Teyateyaneng, Mafeteng and Hlotse, However; about three-fourths of the population is rural areas (Matlosa, 1999).

Figure 1.1 shows the different cities in Lesotho with Maseru highlighted as is the capital city of Lesotho.

Figure 1.1: Map of Lesotho

https://www.worldatlas.com/webimage/countrys/africa/ls.htm

Figure 1.2 shows the most populated cities in Lesotho, with Maseru as the capital having the highest population of 37.66% followed by Teyateyaneng in the Berea region being 12.95% of the population.

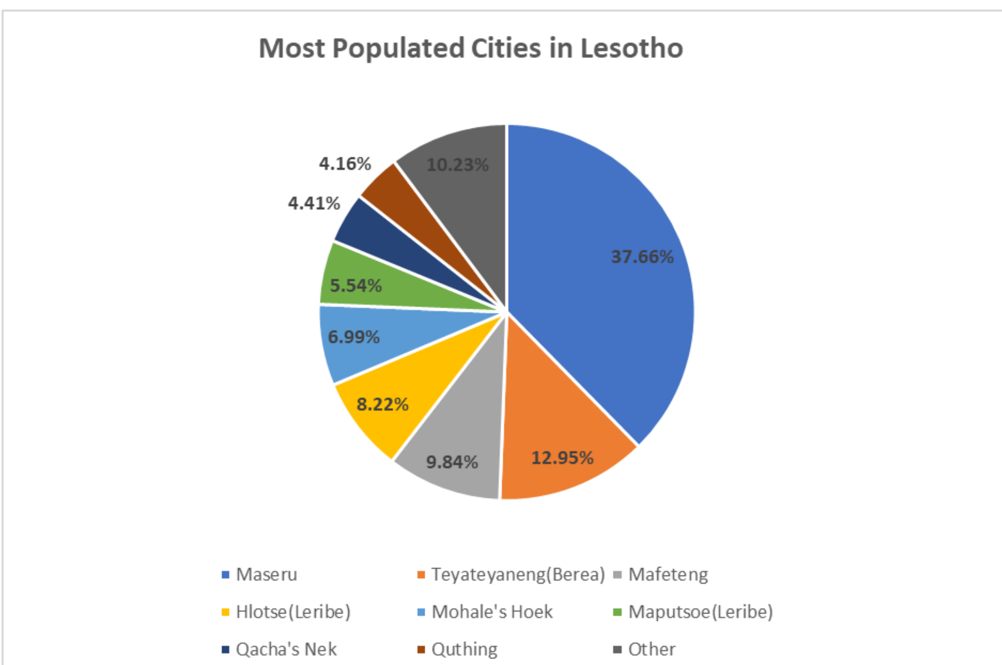

Figure 1.2:Most populated cities in Lesotho

https://www.worldatlas.com/articles/the-biggest-cities-in-lesotho.html

Lesotho is characterized as a lower-middle-income country. The urbanization rate of Lesotho per year is 3.5%. “Almost 40% of the whole population lives underneath the international poverty line which is $1.25 united states dollars per day”(Sumner, 2012).

The country’s birth rate is 28.7 per 1000 people with a total fertility rate of 3.26 births per woman and 51.38% of the whole population is female. Lesotho’s population growth rate is around 1.31% annually. If this trend continues, according to the current population pyramid, the population of Lesotho in the year 2050 will rise to 3.0 million and in the year 2100 to 3,6 million. The life expectancy in Lesotho is 52.5 years, which is low in the international comparison as it ranks number 190 in the world(Worldometers, 2019). The major driver of the huge reduction in life expectancy in Lesotho is HIV and TB co-infection (Gellman, 2000). Swaziland holds the highest HIV prevalence in the whole world followed by Lesotho. TB is the most usual cause of death in people who are HIV-infected.

In Lesotho, 320 thousand people were living with HIV in 2017, 60% children were on antiretroviral treatment and the HIV prevalence was 23.8% (Faturiyele et al., 2018). This outbreak may be the cause of, lesser life expectancy, increased infant mortality rate and increased child mortality rate, increased death rate, changes in sex and age distribution in the population plus a declining population growth.

Lesotho is a country in the sub-Saharan region which is characterized by high child and infant mortality. From the 1970s to 2017, the child mortality rate has reduced in Lesotho from 177 per 1000 live births to 85 per 1 000 live births. In recent years both Infant and child mortality rate has been changing (Satti et al., 2012). Using the data chosen from the 2009 Lesotho Demographic and Health Survey, infant mortality in Lesotho was approximated to be 91 deaths per 1000 births and child mortality was 117 deaths per 1000 births (MOHSW, 2010). The child mortality rate was 85 deaths per 1,000 live births in 2017 while the infant mortality rate was 59 deaths per 1,000 live births. This means that out of 12 children, one dies prior to turning five in Lesotho, and about two-thirds of these deaths happen at the stage of infancy. The LDHS of 2009 showed a rise in the mortality rate of a child for years 2005 to 2009. In the years 2001 up till 2004, the child mortality rate was 113 deaths per 1 000 live births, but from the year 2005 to 2009, the rate became bigger and was 117 deaths per 1 000 live births; This number reduced to 85 in year 2014. The infant mortality dropped too from 91 deaths in 2009 to 59 deaths per 1,000 live births in year 2014. This rate was 74 per 1 000 live births in 2001 (LDHS, 2014).

Lesotho's objective was to decrease by two-thirds the child and infant mortality rate in the interest of the MDG goals, which would have produced a child mortality rate of 37 per 1000 live births and mortality rate for infants to 27 per live births (Ahmad et al., 2000). Lesotho achieved important progress in both the child and infant mortality but did not achieve this target. The only target which Lesotho had set in the year 2000 and substantially achieved was its target of 100% immunization against measles for one-year-olds. As of 2015, the country immunized 90% of its one-year-olds against measles. Lesotho failed to meet MDG 4: Reduce Child Mortality by two-thirds (Hilliard, 2016). Lesotho is amongst the high child-mortality countries at the sub-Saharan Africa. In this study Logistic regression and survey logistic regression will be used to model factors affecting child mortality in Lesotho.

1.3 Health issues, Health Budget and Poverty Profile

The country of Lesotho is facing certain major problems such as high unemployment, TB and HIV/AIDS prevalence rate combined with poverty and gender inequality. These are poor health indices in the country and they affect the infant, children and women in Lesotho(Hassan, 2002). With 57% people living under the national poverty line; Lesotho is one of the low human development countries. In 2014 the HIV prevalence estimated amongst adults ages 15–49 was 24.6 percent. Lesotho's HIV prevalence rate is four times more than the average in the Sub-Saharan Africa (Greg, 2017).

There are 286 health facilities in Lesotho with 20 general district hospitals, 265 primary health care centers and a tertiary hospital in Maseru. Maseru has the highest government funding in the health sector more than five times Leribe; the second-place district. The proportion of doctors to population and nurse-midwives is beneath the WHO Regional Office for Lesotho; having ratio of doctors to population as 0.9 per 10,000 and the ratio of nurse-midwives to population is 10.2 per 10,000 (Unicef, 2017).

Lesotho health sector is funded by the Government, having a minor amount of funding by donors (Akin et al., 1987). The health sector was given the second highest allocation of national budget following education. Regardless, Lesotho's investment in the health sector is lowest compared to neighboring countries. The budget allocations to the health sector have declined significantly in Lesotho while 90 per cent of the health budget is on recurrent expenditure. This shows possibly critical inefficiencies in using the resources like salaries and no funds to create new health centers in remote regions. The budget allocations to child health programs are very low as well and this amount is not enough to fulfill the key challenges of child and infant mortality (Unicef, 2017).

1.4 The importance of the study

Lesotho is amongst one of the highest under-five mortalities in the SSA. "It is important to study under-five mortality because this is a component that's strongly associated with the well-being of a population."(Honwana & Melesse, 2017). The under-five mortality is therefore a useful indicator of health status and the standard of living in a population hence it is importance to analyze it. It also helps to evaluate the effectiveness of intervention of disease-oriented programs that can be introduced for greater impact, which is to decrease child mortality and any formulation of health strategies of the country. A study done by (Pritchard & Keen, 2016) found that there were positive significant correlations between higher child mortality and relative poverty measures in all regions considered. In relation to poverty the study can help in making more potential decisions (Woldemicael, 1999). Analyzing the under-five mortality will help government educate poor families and what put measures in place for households experiencing poverty. "As part of the MDG, targets were set during in 2000, this was the target to decrease child mortality by two-thirds between 2000 and 2015 and it failed with slow progress."(Fotso et al., 2007)

1.5 Objectives

This paper intends to evaluate the factors that affect child mortality. The research paper will look at the relationship between the individual, demographic factors and socio-economic status on child mortality. The key objective of the study is to identify factors that impact child mortality in Lesotho and how they influence the child survival.

- The study looks to define the causes of child mortality in Lesotho.
- The study will model child mortality through logistic regression, as a function of different factors (Wealth index, Mother's level of education, place of residence, sex of child etc.). Therefore, estimating the models using the data.
- Determine how the personal characteristics of the mother and living conditions contribute to the child's survival or death.

The study focuses on the demographic, environmental, and socioeconomic factors that influence child mortality in Lesotho in the periods 2005-2009 and 2010-2014. This will help to know which key sectors the government must give attention to, in order to achieve a decrease in child mortality.

1.6 Data and Methods

1.6.1 Description of the Data

The data used in this study is from the Lesotho Demographic and Health Survey 2009 and 2014. This data is the latest data that exists for the LDHS and the study looks at the latest data available. The objective of 2009 and 2014 LDHS was to provide up to date information on key indicators needed to keep track of progress in Lesotho. The DHS survey authorized access to all the data needed to do the analysis of this study. DHS program stands for Demographic and Health Survey program and was developed and supported by the American People through USAID. It is now the largest and longest enduring survey program of its kind (www.usaid.gov.). It is a survey that collects information on mortality among children and on health status, health indicator. Since 1984, this program has conducted over 300 surveys in over 90 countries, in over a million interviews.

DHS program was established by US agency in 1984 and was implemented in 5-year phases. Years 2015 to 2019 is another phase to be implemented. DHS surveys collect information on child health, fertility, survival, maternal health, contraceptive use, HIV/AIDS, child mortality etc. LDHS data was obtained on request on https://www.dhsprogram.com/ in this study. LDHS sample was selected using stratified, two-stage cluster design. The first stage clusters were selected and in the second stage households were selected from each cluster which yielded the total sample size for each year. The sample size for this data was 3138 for the 2014 dataset and 3999 for the 2009 dataset.

1.6.2 Methodology

Exploratory data analysis will be done to summarize the main characteristics of the data. The statistical models used to explain the relationship between outcome and other variables are listed below:

- Logistic Regression Models,
- Survey Logistic Regression Models,
- Generalized Linear Mixed Models and
- Generalized Additive Models

Statistical Analysis System (SAS) 9.4 are used to fit these statistical models. The results are then interpreted .

1.7 Outline of this study

This study is structured into different chapters. This one gave some background about Lesotho, child mortality, objective of the study, and previous literature on under five child mortality by reviewing papers that have been completed in this area.

A review of the previous research, professional literature and theory is given in Chapter 2. Chapter 3 focuses on the exploratory data analysis which analyzes the data sets to summarize and visualize their main characteristics. Chapter 4,5 and 6 considers different methods to find factors relating to child mortality, it first looks at the logistic regression under SRS. Then the SLR, a model under a generalized linear model framework was used to find the factors related to under-five child mortality to account for the sampling designs complexity. To ease the normality assumptions and the linearity assumption in the parametric models, the semi-parametric generalized additive model, was lastly used for the data. Chapter 7 gives the conclusions and implications drawn from the results found. It also highlights the limitations to the study.

Chapter 2

2 Literature Review

"The rate of child mortality is defined as the number of deaths amongst children under five years old per 1000 live births" (Nannan et al., 2010). These rates can vary depending on the place of residence, mothers education, and level of income as well as other factors (Mosley & Chen, 1984). In the first week of birth the risk of a child death is higher hence safe effective early nutritional care is important in order to stop such deaths. Above 50 percent of the child deaths below the age of five is due to conditions that could be prevented with an access to affordable interventions (WHO & Bank, 2012). This chapter will review some studies that are related to the current study.

Above 10 million children die every year and fifty percent of the deaths happen in just 6 of the countries (Black et al., 2003). Since the implementation of MDG goals, there's been a reduction in the rate of children dying from 12.7 million to 6 million. Despite the growth in the population this has resulted in the global rate of child mortality to reduce from 90 to 43 deaths per 1 000 live births (Liu et al., 2016). More than two fifths of the world deaths take place in SSA. About 7 million children under five years died in 2011 globally with Sub-Saharan African region contributing most to child mortality (Liu et al., 2016). The annual child mortality rate of reduction for the sub-Saharan Africa was more than five times between the years 2005 and 2013 relative to the between years 1990 and 1995. The progress of this was a result of vaccinations taking place (McAllister et al., 2019). Despite the progress made in the sub Saharan Africa, it still contains the highest rates of child mortality having an under 5 mortality rate of 98 deaths per 1000 live births which are high compared to developed regions (Khodaee et al., 2015).

A study done by Ettarh and Kimani in 2012 investigated the determining factors of under 5 mortality in urban and rural areas in Kenya. The Multivariate analysis was used to compare the key risk factors in urban and rural areas in Kenya. Kenya is a country in the sub-Saharan Africa regions as well and had some concerns in achieving the MDG 4 target. This study used data from the 2008-2009 Kenya DHS to determine under-five mortality in urban and rural parts. The research found deaths among under-five children to occur more often in rural areas for mothers age below 21 compared to mothers of the same age in urban areas. The under-five mortality is affected by insufficient breastfeeding, young mothers, poor household and place of residence. A few studies have shown important determinants of under-five mortality to be age of the mother and education of the mother in developing countries. The rate of mortality is higher among mothers that are less educated relative to higher educated mothers. The maternal education is important because it increases a mother's knowledge. The wealth index of the household was a significant factor where households in the rural areas with less wealth were more likely to have child deaths below age five in comparison to rich households (Ettarh et al., 2012). This results to an effective understanding and using the available information and resources for the child survival.

Gebretsadik and Gabreyohanne (2016) performed a study to find the determining factors of child mortality under 5 years in high mortality areas in Ethiopia using the data from Ethiopian DHS in the year 2011. This study was modelled by fitting a Cox proportional regression model to find the factors that affect under-five mortality in the regions between the years 2006 and 2011. 2097 live births were recorded in the five regions of Ethiopia and 366 deaths before five years were reported. In the model fitted, factors that were found to be significant and associated with under 5 mortality were the type of birth, preceding birth interval, the breastfeeding status, family size, the mother's income and the source of water that they drink. The children who breastfed were 25.5%, with p-value equals 0.045, less likely to have died before the age of five than those who were not breastfed. Children who were born following a birth interval of 2 to 3 years and over 3 years were significantly unlikely to die before turning five compared to children who were born within two years. The low probability of child mortality is proportional to increased birth interval time. Therefore, mothers should be waiting for some time after a birth before they conceive a child again (Gebretsadik & Gabreyohannes, 2016).

Motsima studied the risk factors connected to mortality below five years in Lesotho in 2009. A logistic regression was fitted to model the covariates. The study results found female children under-five were significantly 38% unlikely to die in comparison to male children. Children with mothers that stayed in the region Quthing, Qacha's Nek and Thaba Tseka were more likely to die than other regions. A study done suggested that the region, sex of child, breastfeeding and marital status were significant predictors of the under-five mortality (Motsima, 2016).

A study by Coovadia et al., (2007) had the objective of reviewing the data that is available related to child mortality in Africa and its association with the HIV infections status of a mother and child. The research showed that survival of the child is influenced by the HIV epidemic. The study also revealed that child mortality is strongly connected to the maternal health status. The mortality rate of HIV-negative children of HIV-positive mothers was 166 per 1000 live births, but the mortality rate of HIV-negative children for the HIV-negative mother was 128 per 1000 live births. Nevertheless, there is a need for studies on control strategies. This will give improved overall influence of the HIV epidemic on child mortality (Coovadia et al., 2007).

A study done in 2013 to Model Covariates of Child and Infant Mortality in Malawi used 2010 MDHS data. This study's aim was to examine environmental, bio-demographic, and socio-economic factors associated with child and infant mortality in Malawi. Two methods of analysis were applied to the date: The survival analysis and logistic regression method (Lemani, 2013). Study results display that the length of the prior birth interval plus the mother's status of HIV were significantly related to both with child and infant mortality. Other significant covariates included the order of the birth, mothers' age when child is birthed, the wealth index, the sex of the child and education of the parents (Lemani, 2013).

Another study report done by Van Malderen and two other authors found that the birth order, region and birth interval contributed the most to under 5 mortality inequalities in majority of the countries. The aim was to examine and compare the determining factors of inequality in under-five mortality. This study looked at 13 countries in Africa with different under-five mortality and determines the factors that contribute to countries having a higher mortality rate compared to the others (Van Malderen et al., 2013).

They used the multivariate linear regression and the Gini index to discover the contributing determinants of under-five mortality as well as wealth-connected inequality in under 5 mortality in thirteen African countries, including Lesotho. In three of the countries the authors found regional differences to be a significant factor to the inequality. Two other countries had the mother's education as the determinant contributing majority to the overall inequality. Lesotho had unskilled birth attendance and mother's age at birth as one of the important determinants of high under-five mortality. Overall, father's job, mother's education and wealth of household caused more to the inequality, though the most important determining factors were different across countries. This type of study helps in prioritizing interventions aiming to improve the child equity and survival (Van Malderen et al., 2013).

Another study by (Hobcraft et al., 1984) had found similar results when assessing socio economic factors in child mortality for 28 developing countries. Also using a multivariate approach, he also found that child mortality is most strongly associated with only three of the considered variables: husband's job, education and mother's education. In Asian countries, the levels of mother's education are strongly related to mortality in children under five years. The child mortality for the few African countries was relatively strongly associated with education as well as husband's occupation (Hobcraft et al., 1984).

A lot of other studies are available (Consultation, 2012; Sastry, 1997; Uddin & Hossain, 2008), that use the conceptual framework managing the selecting of the analysis variables, which is the proximate determinants model of child health, originally defined by (Mosley & Chen, 1984). Mosley and Chen argued that social science studies should pay attention to the effect of socioeconomic and cultural factors on child mortality. Mosley and Chen portrayed an analytical framework of determining factors of child survival. In their framework, they classify between proximate and socio-economic determining factors of child mortality.

The study identified five important groups of proximate determinants:

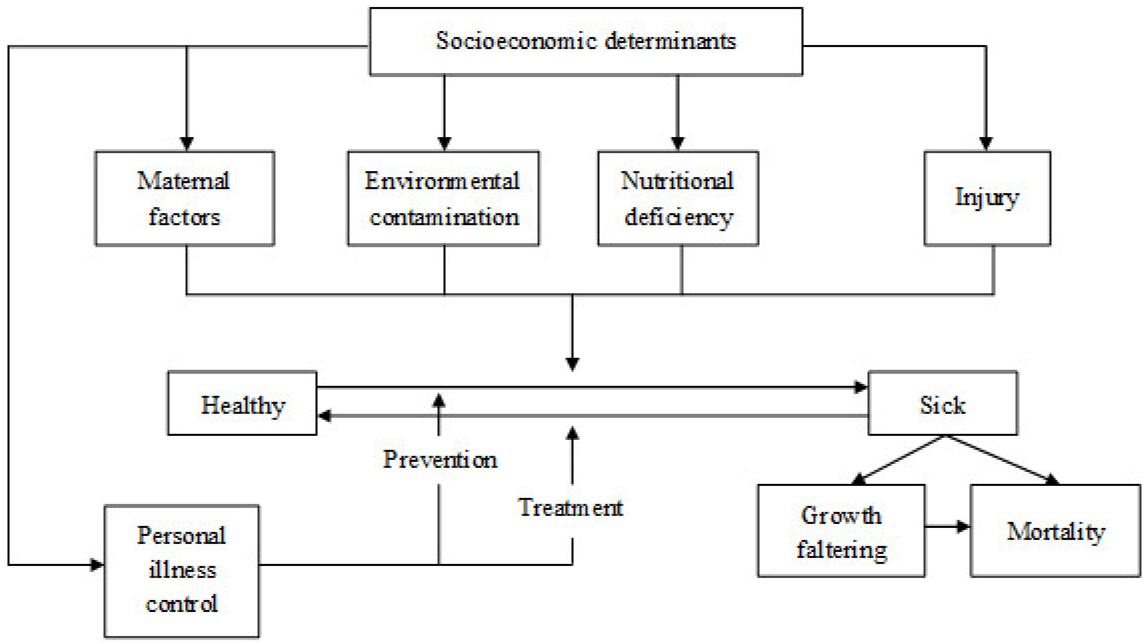

Figure 2.1: Groups of proximate determinants

Source: Adapted from Mosley and Chen (1984).

This study found that the father's education level usually strongly relates to the income of household. They suggested that the father's education may affect the preference on choice of goods. When a more educated father is married to a less educated mother, there is likely to be a more significant effect for child survival, a decrease in child mortality. Mosley and Chen stated that sex of child, birth interval, age of mother and birth order are significant factors affecting child mortality (Mosley & Chen, 1984) .

The study conducted by (Kaundjua, 2013) using the Mosley framework also showed that birth interval, sex of household head, birth weight, toilet facility and mother's education are significantly connected with the risk of dying prior to turning 5.

Chapter 3

3 Exploratory Data Analysis

3.1 Introduction

Exploratory data analysis is helpful to understand the data in detail before the modeling the data. This chapter will summarize main characteristics for the data and will give descriptive statistics such as frequency distributions displayed to describe some of the variables.

3.2 Study Variables

3.2.1 Response Variable

The dependent variable in this study is child survival status that is a binary variable presenting the status of a child being alive or child not alive. The response variable is equal to “zero” if the child is not alive at the time of the survey and “one” if the child is alive.

3.2.2 Exploratory Variables

This study takes into consideration 13 variables including the mothers age of the respondent which was chosen based on the literature. The covariates in the study are geographic, demographic and socio-economic influences. These will be region, age of child, mother’s work status, sex of child, number of children living, mother’s age, mother educational attainment, wealth index, marital status, weight of child when born, type of place of residence, birth order number and breast feeding.

The lists of the explanatory or predictor variables in this study are indicated in Table 3.1. with the codes and descriptions of each variable.

Table 3.1: Description of study variables.

Code	Variables	Description
B4	Sex of child	Male (1), Female (2)
V012	Mother’s age	15 years - 49 years
BORDA	Birth order number	First birth (1), 2-4 births (2), > 4 births (3)

V404	Currently breastfeeding	No (0), Yes (1)
V501AA	Marital status	Not Married (1), Married (2)
M18A	Weight of the child at birth	Large (1), Average (2), Small (3)
B8	Age of child	0, 1, 2, 3, 4
V102	Place of residence	Urban (1), Rural (2)
V101	Region	Botha-bothe (1), Leribe (2), Berea (3), Maseru (4), Mafeteng (5), Mohale's hoek (6), Quthing (7), Qacha's-nek (8), Mokhotlong (9), Thaba Tseka (10)
V190A	Wealth index	Poor (1), Middle (2), Rich (3)
V149A	Education attainment	No Education (1), Up to Secondary Education (2), Higher Education (3)
V714	Mothers currently working	Unemployed (0), Employed (1)
V218A	Number of children living	<2 children (1), 2-4 children (2), > 4 children (3)

3.3 Preliminary Analysis

The goal of this research is finding factors that are associated with under-five years mortality in Lesotho. In order to perform this analysis, certain characteristics need to be examined especially mother's occupation, age, wealth index, birth order number and mother's level of education etc. The mother's age is continuous while the rest of the variables are categorical variables. We will first look at the analysis of the frequency tables which were obtained.

Table 3.2. shows the results for the year 2009, with the sample consisting of n=3999 respondents. The result shows that child sex was almost equally distributed with females accounted for 50.28% and males accounted for 49.71% of the sample thus the 1:1 sex ratio is closely exhibited in the sample. We also see that more respondents were not breastfed and accounted for 53.80% (n=2151) of the sample, and that the sample had 21.10% (n=844) of the respondents that were not married. From the results we can also see that 14.33% (n=573), 66.57% (n=2662) and 19.85% (n=794) of the weight of child at birth were small, average and large respectively. A child is considered small if weight is less than 2.5kg, average if weigh is between 2.5-4.0kg and large if greater than 4.0kg.(Agbozo et al., 2016)

A sample of 672 (16.80%) of the individuals were from urban area, which was three times lower than those in rural areas. More than half of the sample was poor when it comes to the wealth index, 51.8% (n=2071) was poor, 18.6% in the middle and 29.60% was rich. We can see that majority of the mothers were unemployed (64.30%) and most were educated up to secondary school (94.60%). The sample had almost 50% of the children being between second and fourth births with 48.30% and 37.40% are first-borns. More than half of the sample had 2-4 children living in the household, accounting for 51.60%(n=2063) out of the whole sample.

Table 3.2: Summary of selected characteristics of children (LDHS 2009).

Covariates	**Characteristics**	**Frequency**	**Percent (%)**
Sex of the child	Male	1988	49.71%
	Female	2011	50.28%
Age of child	0 years	1162	29.06%
	1 years	840	21.00%
	2 years	670	16.75%
	3 years	657	16.43%
	4 years	670	16.75%
Mother's current age	Mother's current age	3999	100.00%
Weight of child at birth	Large	794	19.85%
	Average	2662	66.57%
	Small	573	14.33%
Currently breastfeeding	No	2151	53.80%
	Yes	1848	46.20%
Marital Status	Not Married	844	21.10%
	Married	3155	78.89%
Wealth Index	Poor	2071	51.8%
	Middle	745	18.6%
	Rich	1183	29.6%
Education Attainment	No Education	90	2.30%
	Up to Secondary Education	3783	94.6%
	Higher Education	126	3.20%
Mothers currently working	Unemployed	2571	64.30%
	Employed	1428	35.70%

Type of place of residence	Urban	672	16.80%
	Rural	3327	83.20%
Region	Butha-Bothe	357	8.90%
	Leribe	427	10.70%
	Berea	358	9.00%
	Maseru	498	12.50%
	Mafeteng	357	8.90%
	Mohale's Hoek	362	9.10%
	Quthing	332	8.30%
	Qacha's-Nek	317	7.90%
	Mokhotlong	482	12.10%
	Thaba-Tseka	509	12.70%
Birth Order Number	First Birth	1494	37.40%
	2-4 Births	1932	48.30%
	>4 Births	573	14.30%
Number of children living	<2 children	1398	35.00%
	2-4 children	2063	51.60%
	>4 children	538	13.50%
Total		3999	100.00%

The results in Table 3.3. show that the sample consisted of n=3138 respondents in the year 2014. The result shows that child sex was almost equally distributed with males accounted for 49.60% and females accounted for 50.40% of the sample thus the 1:1 sex ratio is closely exhibited in the sample. We also see that more respondents were not breastfed and accounted for 57.70% (n=1812) of the sample, and that the sample had 21.54% (n=676) of the respondents that were not married. From the results we can see that 14.72% (n=462), 66.32% (n=2081) and 18.96% (n=595) of the weight of child at birth was small, average and large respectively.

With regards to the type of residence, there were 25.05% (n=786) individuals from urban area, which was three times lower than those in rural areas. Almost half of the sample was poor when it comes to the wealth index, 46.56% (n=1461) was poor, 20.27% in the middle and 33.17% was rich.

The data also shows that majority of the mothers were unemployed (70.01), and most were educated up to secondary school (92.26%). The sample had 49.43% of the children being between second and fourth births and 39.01 as first born. 53.60% of the sample had 2-4 children living in the household, accounting for n=1682 out of the whole sample, with 1160 of the sample have less than two children in household.

Table 3.3: Summary of selected characteristics of children (LDHS 2014).

Covariates	Characteristics	Frequency	Percent (%)
Sex of the child	Male	1555	49.60%
	Female	1583	50.40%
Current age of child	0 years	875	27.88%
	1 years	680	21.67%
	2 years	582	18.55%
	3 years	492	15.68%
	4 years	509	16.22%
Mother's current age	Mother's current age	3138	100.00%
Weight of child at birth	Large	595	18.96%
	Average	2081	66.32%
	Small	462	14.72%
Currently breastfeeding	No	1812	57.70%
	Yes	1326	42.30%
Marital Status	Not Married	676	21.54%
	Married	2462	78.46%
Wealth Index	Poor	1461	46.56%
	Middle	636	20.27%
	Rich	1041	33.17%
Education Attainment	No Education	36	1.15%
	Up to Secondary Education	2895	92.26%
	Higher Education	207	6.60%
Mothers currently working	Unemployed	2197	70.01%
	Employed	941	29.99%
Type of place of residence	Urban	786	25.05%

	Rural	2352	74.95%
Region	Botha-Bothe	298	9.50%
	Leribe	351	11.19%
	Berea	323	10.29%
	Maseru	391	12.46%
	Mafeteng	272	8.67%
	Mohale's Hoek	307	9.78%
	Quthing	296	9.43%
	Qacha's-Nek	236	7.52%
	Mokhotlong	336	10.71%
	Thaba Tseka	328	10.45%
Birth Order Number	First Birth	1224	39.01%
	2-4 Births	1551	49.43%
	>4 Births	363	11.57%
Number of children living	<2 children	1160	37.00%
	2-4 children	1682	53.60%
	>4 children	296	9.40%
Total		3138	100.00%

From Table 3.2. and Figure 3.1 we can see that out of all the regions in 2009, the majority of the sample was from Thaba-Tseka with 12.70% (n=509) followed by Maseru 12.50% (n=498), Mokhotlong 12.10% (n=482) and Leribe 10.70% (n=427). For 2014, Table 3.3. and Figure 3.1 show that the majority of the sample was from Maseru (12.46% ,n=391) followed by Leribe (11.19% , n=351), Mokhotlong (10.71% , n=336) and Thaba Tseka (10.45% , n=328). Figure 3.2 shows the percentage of children alive compared to child deaths for the years 2009 and 2014. In a sample of 3999 in 2009, n= 393 children died accounting for 9.83% of the sample; while in 2014, n=223 (7.11%) children died out of sample 3138. Figure 3.3 shows wealth index of the mother for the years 2009 and 2014. It shows that majority of the mothers are poor in Lesotho (52% and 47%). An assessment for poverty revealed that poverty has decreased over time in Lesotho for the past 15 years, but it remains widespread with almost half of the population surviving with poverty and 75% of the population is poor (Bank, 2019).

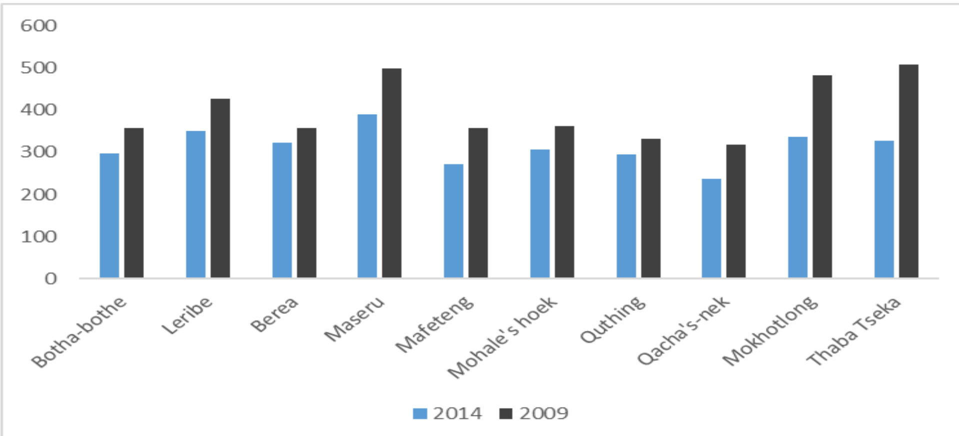

Figure 3.1: Sample distribution of children for each region in Lesotho for the two periods.

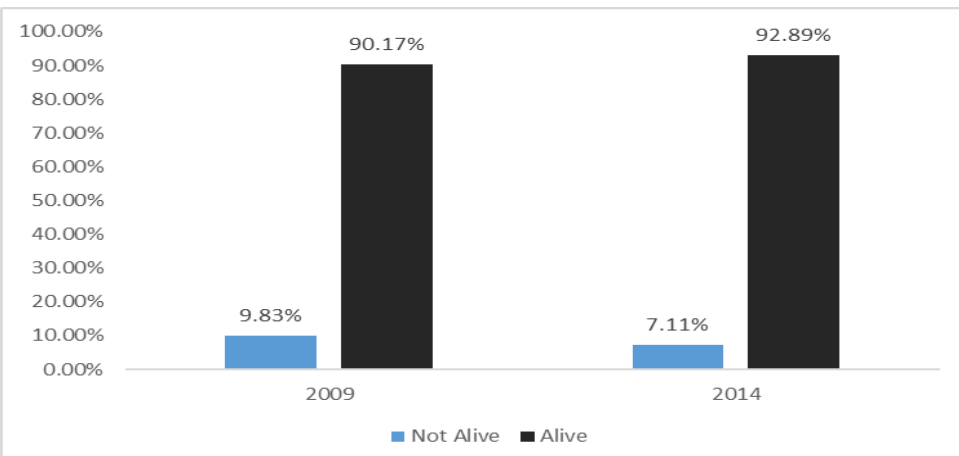

Figure 3.2: Percentage of child survivals for 2009 and 2014 in Lesotho

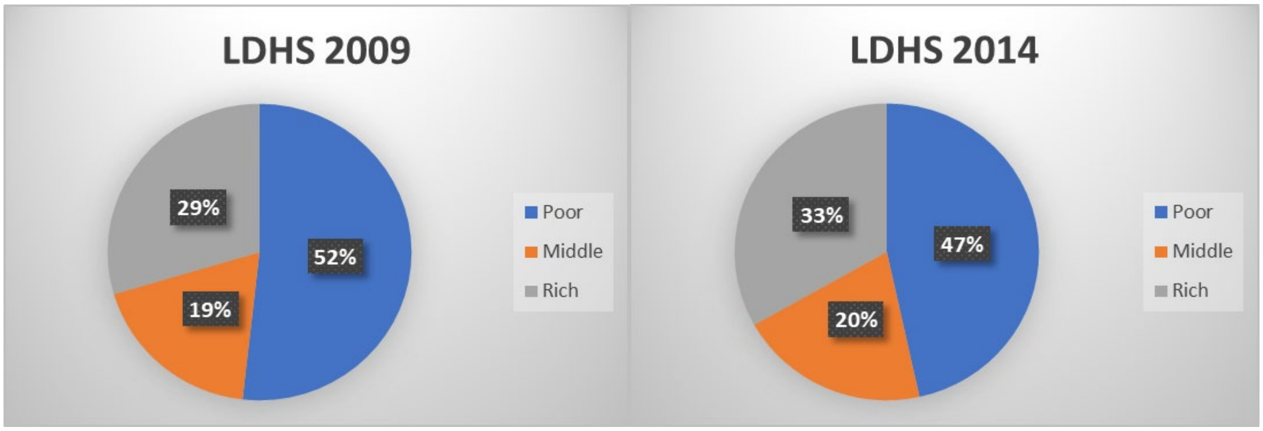

Figure 3.3: Wealth Index of Mother in Lesotho (2009 and 2014).

3.4 Summary

The exploratory data analysis is important in helping us get a preliminary view of the trends and patterns of the data before using model-based approaches. The results are easier to interpret because the EDA analysis results can be seen analytically and visually. Assessing variables by using descriptive measures such as central tendency as well as graphical methods helps to summarize the sample and give an overview of the whole data.

By using the frequency tables with the summarized data , we can see that overall, for both 2009 and 2014 samples, it has been observed that most of the respondent's children were under the age of one and that respondents who were married were above 78%(Table 3.3). There were slightly more females than males. Most of the people that responded lived in rural parts also 46.56% of the samples were poor. The number of children in a household is mostly 2-4 children in each household with weight of child being average for most children in the samples. Children birth order number between 2 and 4 had the majority proportion of the samples.

An attempt to present the description of the data was made in this chapter. The next chapter applies some statistical methods to study the relationship between independent and response variables

Chapter 4

4 Generalized Linear Models

4.1 Introduction

The response variable in this study is survival status of a child which is a dichotomous variable showing the status: of a child alive or not. This binary outcome variable follows a Bernoulli distribution which is a member of the exponential family (Tutz, 2011).

The generalized linear model widens the general linear model in such a way that the model permits for the response variable to take on a distribution that is unlike the normal distribution (Hardin et al., 2007). The generalized linear model includes statistical models like: logistic regression model, loglinear models for count data and categorical data analysis, complementary log-log models for interval-censored data, as well as numerous other statistical models through its very general model formulation (Fahrmeir & Tutz, 2013). In comparison to these statistical models, the logistic regression is the most common technique used in health science (Tetrault et al., 2008). The methods utilized in a logistic regression follows similar principles used in linear regression (Dobson & Barnett, 2008). To be able to make valid statistical inference all variables which possibly affects the child death will be assumed to have fixed effects. Subsequently, we fit the Generalized Linear Model to the data. Below we look at the theory of the GLM.

4.2 The Generalized Linear Model

The General Linear Model (GLM) is mathematically the same as the multiple regression analysis but stresses its suitability for both qualitative and quantitative variables. The multiple regression analysis encompasses all linear models including linear regression model for a continuous response variable and continuous predictor or continuous response or categorical predictor (Cohen, 1968). The model has a dependent variable that is normally distributed with mean, μ_i, a which is a function of $X_i^T \beta$ and constant variance σ^2 where x_i has known covariates and β comprises of the coefficients that will be estimated.

The generalized linear model makes use of the linear functional relationship to explain how a response variable depends on measured explanatory variables. This response variable, y_i, is presumed to come from an exponential family of distribution with mean μ_i, which is a function of ${x_i}^T\beta$ (McCullagh & Nelder, 1989). The GLM generalizes linear models by allowing the error to take a distribution except a normal distribution. The general linear model is given by:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \epsilon \tag{4.1}$$

where, $\boldsymbol{Y}$ is the response variable, $\boldsymbol{X}$ refers to the design matrix of covariates, β stands for the vector of regression coefficients and ϵ stands for error vector.

The GLM assumes that the errors should be independent but not necessarily normally distributed. This permits the distribution to be some distribution from exponential family (Dobson & Barnett, 2008). "It permits the linear model to be connected to the response variable through a link function, letting the size of the variance of each measurement to be a function of its predicted value" (Hutcheson & Sofroniou, 1999). We do not necessarily model the mean directly , but we model some function of the mean $g(\mu)$; This can be specified as

$$g(\mu) = \eta = \boldsymbol{X}\,\beta \tag{4.2}$$

In the above, $\eta = \boldsymbol{X}\beta$ is the linear predictor part of the model and g(.) is the link function (Müller & Stadtmüller, 2005). Let us first look at exponential family of distributions.

4.2.1 Exponential Family of Distributions

This family of distributions consists of a lot of distributions which are applicable and valuable for modeling that is practical for instance: Negative Binomial or Poisson for count response variable; Bernoulli , Binomial, Geometric for studying discrete responses; Normal, Beta, Gamma, Inverse Gaussian, as well as exponential for studying continuous responses (Hardin et al., 2007).

One can confirm that a distribution is from the exponential family of distribution if the probability distribution function (pdf) can be expressed as below.

$$f(y_i|\theta_i, \phi) = exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right) \tag{4.3}$$

where, $a(\phi)$ and $b(\theta_i)$ are known functions with ϕ being the scale parameter and θ_i the canonical parameter. $c(y_i, \phi)$ is some function of y_i and ϕ.

4.2.2 Components of the GLMs

There are three components of generalized linear model. The random component that characterizes the response variable and its probability distribution. The systematic component denotes the explanatory variables ($\mathbf{X}_1$, $\mathbf{X}_2$, ... $\mathbf{X}_\mathbf{k}$) in the model, more precisely their linear combination in creating the linear predictor; for example, $\beta_0 + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + ... B_k\mathbf{x}_k$ as we will see in a logistic regression in this analysis. The link function indicates the link between systematic and random components. It shows how the mean response connects to the linear predictor of explanatory variables (Bonnefoix et al., 1996; McCullagh, 2019).

There are several of advantages of the GLMs in comparison to the OLS regression but there also exists limitations. In the GLM, transformation of the response variable is not necessary to have a normal distribution, the models are fitted via MLE and so ideal properties of the estimators. The decision of the link is different from the decision of random component and hence there is more flexibility in modeling; inference tools and model checking e.g., Residuals, Deviance Confidence intervals, Wald and Likelihood ratio tests, Overdispersion apply. The limitations are linear function; it can only have a linear predictor in the systematic component and the responses has to be independent (Lee & Nelder, 2006; Levy, 2012).

4.2.3 Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation is a technique that finds values for the models' parameters. The parameter values maximize the likelihood that the process described by the model produced the data that was seen. A technique such as Fisher's scoring or Newton-Raphson must be utilized to find MLEs for the GLMs. The technique of Maximum Likelihood is advised to be more strong and yields estimators with great statistical properties in comparison to other methods such as the technique of LSE (Hardin et al., 2007; Myung, 2003). Because of advanced computer software and statistical theory, this technique of estimation is known to be the most known technique in applied statistics (Wu, 2005).

4.3 Model Selection and Diagnostics

4.3.1 Model Selection

Selecting the Model is a critical portion of all statistical analysis and is generally the principal to the pursuing science. The selection of the model includes the selecting the best model out of a few competing models. The criteria of selection for the model that are usually used for GLMs are; Bayesian information criterion (BIC) and Akaike information criterion (AIC) (Gayawan & Ipinyomi, 2009). If the maximum likelihood is used to approximate the parameters and the models are non-nested, then the Bayes information criterion or the Akaike information criterion may be used to compare the models (Vrieze, 2012). The model selection criteria examines the minimum AIC and BIC values and the best model is selected based on this (Wang & Liu, 2006). However, the likelihood ratio test is used when the models are nested.

4.3.1.1 Akaike's Information Criterion (AIC)

The Akaike information criterion is a method based on in-sample fit to approximate the likelihood of a model to forecast the values in the future. The AIC criterion tries to measure how good the model has estimated the data. The Akaike's Information Criterion is a practical statistic to compare the comparative fit of various models (Li & Nyholt, 2001).

The statistic suggested by Akaike (1974) is:

$$\mathrm{AIC} = -2\ln(L) + 2k$$

the number of parameters is given by k, and the maximum log-likelihood is denoted by ln(L). The AIC technique penalizes the loglikelihood for the number of parameters approximated (Akaike, 1974). A model with a minimum AIC in comparison to all the other models is a good model (Arnold, 2010).

4.3.1.2 Bayesian information criterion

The Bayesian information criterion is an alternative method for model selection that calculates the trade-off relating to the model fit and how complex the model is. The BIC is also known as the Schwarz Criterion (SC). This is an alternate method to the AIC for comparison of nested models and was suggested by Schwarz in 1978 (Müller et al., 2013).

The BIC statistic (Stone, 1979) is given by

$$SC = -2\ln(L) + 2\ln(N)k. \tag{4.4}$$

N is the sample size; L is the likelihood; and k are the number of parameters to be estimated. The choices can be reduced before comparing the models. This is done by choosing variables that are incorporated in the model.

The selection of variables that are incorporated in the model is completed by three procedures specifically, stepwise, backward and forward selection. The forward selection initiates with a null model and includes one explanatory variable individually against a certain level of significance α, up till every significant variable is incorporated in the model. The backward selection begins with the model that contains all covariates and drops one insignificant at a time. This is done till every non-significant variable is removed in the model. The stepwise selection performs like the forward selection procedure. The benefit of stepwise selection is that variables existing in the model are given thought to for the omission in the model whenever a new covariate is added in the model (Heinze et al., 2018).

4.3.1.3 Measure of fit

An important stage in statistical analysis is to evaluate the goodness-of-fit of the model. The deviance and Pearson chi-square tests provide large sample tests of the model fit (Hosmer et al., 1997). However, the deviance has an advantage over Pearson's chi-square statistic because for nested models it is additive (Nelder & Baker, 1972).

4.3.2 Model Checking

4.3.2.1 Deviance

The deviance is used to evaluate the fit in the GLMs and also be used in comparing nested models. It looks at measuring the inconsistency between the actual values from the dataset and predicted values from the fitted model. To put differently; it is the disparity of fit between the maximum log-likelihood of the log-likelihood of the model fitted and the saturated model (Spiegelhalter et al., 2002).

Let $\ell(\hat{\mu}, \phi, y)$ denote the log-likelihood of the model that is reduced at the MLE , $\ell(y, \phi, y)$ stand for the log-likelihood estimate of the saturated model. Deviance can be defined as:

$$Deviance = 2[\ell(y, \phi, y) - \ell(\hat{\mu}, \phi, y)] \tag{4.5}$$

The MLEs is given by $\hat{\mu}$, and $\phi = 1$; the scaled deviance is given by the below equation when $\phi \neq 1$:

$$Deviance = \frac{2[\ell(y,\phi,y) - \ell(\hat{\mu},\phi,y)]}{\phi} \tag{4.6}$$

when the ϕ is not known, we can approximate it by:

$$\hat{\phi} = \frac{D}{n-p}$$

the number of variables is p and the number of observations is given by n.

4.3.2.2 Pearson's Chi-square

The alternative to deviance for testing and comparing models is the generalized Pearson chi-square statistic. The score statistic for testing the fitted model against the saturated model is called the Pearson goodness of fit statistic defined as (Smyth, 2003):

$$Pearson's = \chi^2 = \sum_{i=1}^{n} \frac{(y_i - \widehat{\mu_i})^2}{V(\widehat{\mu_i})} \tag{4.7}$$

$V(\widehat{\mu_i})$ is the estimated variance function which is equivalent to $\mu_i(n_i - \mu_i)/n_i$ and $\mu_i = np_i$. This statistic asymptotically follows χ^2 distribution with n-p-1 degrees of freedom. The bigger the value of the χ^2statistic, the worse the model fit (Cochran, 1952).

4.3.2.3 Link Function

The selection of link function is essential. If the link function is unsuitable, then the estimated results will be erroneous and can result to incorrect conclusions. To test how appropriate the link function is, we can refit the model using the linear predictor acquired from the initial model as well as the square of the linear predictor as explanatory variables (Faraway, 2016; Moeti, 2007).The linear predictor will be significant if the link function is suitable and the squared linear predictor term will be insignificant. The initial model may be alternatively estimated with an additional composed variable, in this case, for an appropriate model the extra variable is statistically insignificant (Venables & Ripley, 1999). If the constructed variables are significant, then the link function is unsuitable or crucial factors are excluded in the model (Bolker et al., 2009).

4.4 Background to Logistic regression

The logistic function was established as a model of population growth and was given the name logistic by Pierre Francois Verhulst in the 1830s. It was re-established afresh in 1920 by Pearl and Reed in a research of the population growth of the United States (Reed, 1929). The logistic regression model was established in 1958 by David Cox who was a statistician. To estimate the probability of a binary response variable based on a single or more independent variables the binary logistic regression can be used. This permits us to say that the existence of risk factors increases the probability of the outcome by a certain percentage (Cox, 1958; Cramer, 2003).

The Logistic regression is the suitable analysis when the dependent variable is binary. It is adapted to explain the data and the connection between a single dependent dichotomous variable and a single or more than one independent variables (Bewick et al., 2005; Tranmer & Elliot, 2008). In this we research we focus on the logistic regression model with more than one predictor variable known as multiple logistic regressions. We shall limit ourselves to the use of binary logistic regression, where our response variable is strictly dichotomous depicting whether a child is alive or not.

4.4.1 Logistic Regression Model

To model binary data; a member of the GLMs called the logistic regression model is used. This model describes the probability of a specific event, that is a value between 0 and 1 (Melesse et al., 2016). Assume we have a dichotomous outcome variable Y; $Y = 1$ if the event occurs and $Y = 0$ if the event does not occur. We have a set of p independent variables denoted by the vector $\boldsymbol{x} = (x_1; x_2; \dots ; x_p)$.

It is assumed that Y has a Bernoulli distribution represented as:
$Y = 1$; when the event occurs (π_i)
$Y = 0$; when the event does not occur $(1 - \pi_i)$
Thus, Y_i follows a Bernoulli distribution with $P(Y_i = 1) = \pi_i$ and $P(Y_i = 0) = 1 - \pi_i$
Therefore,

$$E(Y_i) = \pi_i \tag{4.8}$$

$$Var(Y_i) = \pi_i(1 - \pi_i) \tag{4.9}$$

For k explanatory variables and $i = 1, \dots, n$ individuals,

With regards to the odds ratio, the logit model is:

$$logit(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi}. \tag{4.10}$$

The formula that refers to the probability of the outcome is:

$$\pi_i = P(Y_i = 1|X_i = x_i) = \frac{exp(\beta_0+\beta_1 x_{1i}+ \dots +\beta_p x_{pi})}{1+exp(\beta_0+\beta_1 x_{1i}+ \dots +\beta_p x_{pi})} \quad (4.11)$$

This is the probability of the event happening, and the probability of the event not happening is given by $1 - \pi_i$. This the logistic regression model which is a member of the GLM with a logit link. The odds ratio is described as the ratio of the odds event taking place in one group to it happening in another (Cleves et al., 2008). The value of the logit model when the values of predictor variables are equal to zero is the intercept (β_0). This means there are no predictor variables.

The regression coefficient tells us how much the dependent variable is expected to increase the probability of the model if the coefficient is positive or decrease it if the coefficient is negative(Williams, 2016). The probability of the model is strongly influenced by the variable if the coefficient is large, while a very small coefficient indicates that variable has small impact on the probability of the model. A confidence interval for each variable shows the uncertainty in the estimate (Liao, 1994).

A model containing nominal(categorical) variables has a more complex interpretation of the coefficients. Dummy predictor variables are created for every term that involves a nominal variable to predict the effect. The logistic regression uses reference dummy coding to code the predictors for a nominal variable (Long & Freese, 2006). The first category is the baseline and the other coefficients can be interpreted as an increase or decrease in the log-odds ratio over the baseline category (Long & Freese, 2006; Mitchell, 2012).

The objective of logistic regression is obtaining the probability of a certain event knowing the independent variables. Using Logistic regression, we model the LDHS 2014 data to find variables that best fit the data associated to child mortality.

4.4.2 Assumptions of Logistic Regression

1) ASSUMPTION OF APPROPRIATE OUTCOME STRUCTURE

The binary logistic regression needs the response variable to be binary whilst the ordinal logistic regression needs the response variable to be ordinal.

2) ASSUMPTION OF OBSERVATION INDEPENDENCE

For the Logistic regression, the observations need to be independent of each other. So, the observations shouldn't derive from measurements that are repeated.

3) ASSUMPTION OF THE ABSENCE OF MULTICOLLINEARITY

The model should have little to no multicollinearity between the independent variables. The independent variables shouldn't be too highly correlated with one another.

4) ASSUMPTION OF LINEARITY OF INDEPENDENT VARIABLES AND LOG ODDS

The linearity of independent variables and log odds is assumed in logistic regression. The analysis needs the independent variables to be linearly linked to the log odds.

5) ASSUMPTION OF A LARGE SAMPLE SIZE

The model requires a large sample size.

Regression analysis suggests the significant relationships between dependent variable and independent variable. It also shows how strong the effect of multiple independent variables is on a dependent variable (Lani, 2010; Morrow-Howell & Proctor, 1993; Starkweather & Moske, 2011).

4.4.3 Limitations to the Logistic Regression

The Logistic Regression is not the strongest model out there and can easily be outperformed by more complex models. Logistic regression is not useful unless you have identified all the crucial predictor variables (Bursac et al., 2008). In logistic regression, zero assumptions are stated about the distributions of the independent variables, but independent variables shouldn't be highly correlated to each other since it may lead to issues with estimation (Horton & Fitzmaurice, 2004).

Logistic regression works well for predicting nominal outcomes like a child status. It can also predict multinomial outcomes, like rejection or wait listing in a university (Peng et al., 2002). We cannot use the Logistic regression to find how high an influenza patient's fever will increase or decrease, because the scale of measurement is continuous (Hosmer Jr et al., 2013). Therefore, using the logistic regression modelling to find continuous outcomes will not give the best results.

The logarithm of the odds of the dichotomous outcome in the model is calculated as a linear combination of the (Lalonde et al., 2013). The limitation one that when there is non-linear relationship between covariates and log odds we could obtain invalid results (Peterson & Harrell Jr, 1990). The other limitation is that the ordinary logistic regression does not account for how complex the survey design is, which can lead to statistical inference that is invalid (Horton & Fitzmaurice, 2004).

"The Hosmer-Lemeshow goodness of fit test is based on dividing the sample in accordance with their predicted probabilities" (Guffey, 2013). The Hosmer-Lemeshow test has a lower power to identify an incorrect specification if the sample is small (Allison, 2014; Hosmer et al., 1997).

4.4.4 Parameter Estimation

Maximum Likelihood Estimation of Logistic Regression Models

The parameters (β_i) are unknown in the model are estimated applying the maximum likelihood estimation procedure. The statistical assumption of independence of observations applies. In this study, a variety of respondents in the Lesotho DHS data are presumed to be independent of each other. The parameter value whose probability of the data takes its biggest value is the maximum likelihood estimate (Albert & Anderson, 1984; Fienberg & Rinaldo, 2007).

“The maximum likelihood equation is attained from the probability distribution of the predictor variable” (Miranda & Rabe-Hesketh, 2006). Since each y_i represents a binomial count in the ith population, the probability distribution of a binomial response variable $Y_i \sim B(n_i, \pi_i)$ is given by:

$$P(Y_i = y_i) = \left(\frac{n_i!}{y_i!(n_i - y_i)!}\right) {\pi_i}^{y_i}(1 - \pi_i)^{n_i - y_i} \tag{4.12}$$

for $y_i = (0,1,2, \ldots, n_i)$ where ${\pi_i}^{y_i}(1 - \pi_i)^{n_i - y_i}$ is the probability of obtaining y_i successes and $n_i - y_i$ failures with a combinatorial coefficient $\binom{n_i}{y_i}$ which is the number of ways of obtaining y_i successes in n_i trials. The mean and the variance of Y_i are given by:

$$E(Y_i) = n_i \pi_i \tag{4.13}$$

$$V(Y_i) = n_i \pi_i (1 - \pi_i) \tag{4.14}$$

If $\binom{n_i}{y_i}$ is a constant, then we have the likelihood function being of importance and is written as:

$$L(\pi_i) = \prod_{i=1}^{N} {\pi_i}^{y_i}(1 - \pi_i)^{n_i - y_i} \tag{4.15}$$

considering the natural log, we get the log likelihood as:

$$logL(\pi_i) = log \prod_{i=1}^{N} {\pi_i}^{y_i}(1 - \pi_i)^{n_i - y_i} \tag{4.16}$$

$$logL(\pi_i) = \sum_{i=1}^{N} \left(y_i log(\pi_i) + n_i - y_i log(1 - \pi_i)\right) \tag{4.17}$$

There is one parameter connected to each of the K columns of predictor variable in X, plus one β (0), for the intercept. β is the parameter vector, a column vector of length K + 1. The objective of the logistic regression is finding the K +1 not known parameters. In solving the parameter estimates, make the first derivative of log-likelihood with regards to each β is 0, the MLEs for β is attained by making each of the K+1 equation established equals 0 and working out for each $\beta(k)$.

If a solution exists, each solution stipulates a critical point that's a maximum or minimum. If the matrix of second derivatives is negative definite then the critical point will be the maximum, which means that every element on the diagonal of the matrix is below zero. The matrix forms variance-covariance matrix of the parameter estimates. When we differentiate each of the K + 1 equation again in terms of each β we get the variance covariance matrix (Czepiel, 2002; Dorfman & Alf Jr, 1969). The parameters can therefore be estimated by finding score equations and using an iterative method such as the Fisher scoring algorithm or Newton-Raphson method to solve for their values (Pourahmadi, 2000).

Iterative methods that are mostly used include the Newton-Raphson and Fisher scoring technique. Other iterative methods do exist like IRLS and IWLS. The Newton-Raphson method makes use of the standard least squares technique to repeatedly compute the maximum likelihood estimates (Yirga, 2018) and the Fisher scoring technique is equal to the IRLS (Wedel & DeSarbo, 1995).

The PROC LOGISTIC is the most common SAS procedure to fit the logistic regression. It gives Maximum Likelihood Estimation of the logistic regression model, that employs by default the Fisher's scoring technique (Allison, 2012).

4.4.5 Newton-Raphson Method

The logistic regression aims to estimate P+1 variables β that are unknown in this equation:

$$log\left(\frac{\pi_i}{1-\pi_i}\right) = \sum_{\mathrm{j=0}}^{\mathrm{p}} \beta_j x_{ij}. \tag{4.18}$$

By obtaining the set of variables for which the probability of the observed data is greatest; this aim can be achieved with Maximum Likelihood Estimation. The maximum likelihood estimates for β may be obtained by equating P + 1 equations individually in the loglikelihood to zero and solving for β(j). Setting the equations in equation (4.18) to zero results in a system of P + 1 non-linear equations each with P + 1 variables that are not known. The solution to the system is a vector with elements, β_j. After confirming that the matrix of second derivatives is negative definite, and the solution is the global maximum and not a local maximum, then this vector comprises of the parameter estimates for which the observed data will have the greatest probability of occurring. The solution for solving a system of nonlinear equations cannot be derived algebraically. An iterative process is used to numerically estimate the solution.

The most popular technique for working out the systems of nonlinear equations is the Newton-Raphson method. “This method starts with a guess for the solution; it then uses the first two terms of the Taylor polynomial estimated at the initial guess to get another estimate that is closer to the solution” (Czepiel, 2002; Jennrich & Robinson, 1969). This process continues until it converges to a solution.

4.4.6 Parameter Estimation

The technique of maximum likelihood is the theoretical foundation for parameter estimating in GLMs, where the mean response is connected to the linear predictors by the link function, with g(μ) = Xβ (Collett, 2002). The likelihood function is:

$$L(y;\theta) = \prod_{i=1}^{n} exp\left(\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i,\phi)\right) \tag{4.19}$$

$$= exp\left(\sum_{i=1}^{n}\left(\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i,\phi)\right)\right) \tag{4.20}$$

The log-likelihood is written as

$$\ell(y;\theta) = \log(y;\theta) = \sum_{i=1}^{n}\left(\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right) \tag{4.21}$$

Because y_i , $i = 1,\ldots,n$, are independent, the joint log-likelihood function is

$$\ell(\beta, y) = \sum_{i=1}^{n} \ell_i. \tag{4.22}$$

with

$$\ell_i = \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi). \tag{4.23}$$

The parameter estimates are estimated by differentiating the log-likelihood function with regards to βj, equating derivatives to 0, and after solving the system of equations at the same time for βj. That is

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell_i}{\partial \beta_j}. \tag{4.24}$$

$j = 0, 1, 2, \dots, p$

Here p+1 is the number of parameters.

Use the chain rule of differentiation, $\frac{\partial \ell}{\partial \beta_j}$ is computed as

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}. \tag{4.25}$$

From equation 4.25., the first factor becomes

$$\frac{\partial \ell_i}{\partial \theta_i} = \frac{y_i - b\prime(\theta_i)}{a_i(\phi)} = \frac{y_i - \mu_i}{a_i(\phi)}. \tag{4.26}$$

Because $\mu_i = E(y_i) = b'(\theta_i)$

and $V(y_i) = a_i(\phi)V(\mu_i)$.

Factor number two is

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{b\prime\prime(\theta_i)} = \frac{a_i(\phi)}{Var(y_i)}. \tag{4.27}$$

The third factor is dependent on the link function. The linear predictor is

$$\eta_i = g(\mu_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{ip}. \tag{4.28}$$

and

$$\frac{\partial \mu_i}{\partial \eta_i} = b''(\theta_i). \tag{4.29}$$

And the last factor

$$\frac{\partial \eta_i}{\partial \beta_j} = x_{ij}. \tag{4.30}$$

Where, x_{ij} is the jth element of covariates vector x_i for the ith observation.

Substituting all the factors

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{\partial \ell_i}{\partial \theta_i}\frac{\partial \theta_i}{\partial \mu_i}\frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j}. \tag{4.31}$$

We get

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{y_i - \mu_i}{a_i(\phi)}\,\frac{a_i(\phi)}{Var(y_i)}\,b''(\theta_i)x_{ij}. \tag{4.32}$$

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{y_i - \mu_i}{Var(y_i)}\,b''(\theta_i)x_{ij}. \tag{4.33}$$

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{y_i - \mu_i}{a_i(\phi)}\,x_{ij}. \tag{4.34}$$

Therefore, the systems of equations to be solved for β_j's is given by

$$\frac{\partial \ell_i}{\partial \beta_j} = \sum_{i=1}^{n}\frac{y_i - \mu_i}{a_i(\phi)}\,x_{ij} = 0 \tag{4.35}$$

The equations are solved repeatedly. A starting solution of the equations given by $\hat{\beta}^{(0)}$ is approximated and then revised up to the convergence of iterative algorithm to the solution $\hat{\beta}$, named the maximum likelihood estimate of β. Many statistical packages have iterative algorithms for solving , such as GenStat, STATA, SAS etc. (Jennrich & Sampson, 1976). The system of equations can be worked out repeatedly by making use of the Fisher's scoring or Newton-Raphson algorithms for MLE (McCullagh & Nelder, 1989). These two techniques are the most widely utilized algorithms for MLE.

The Fisher's scoring technique is equal to the iterative reweighted least-squares (IWLS) (Green, 1984; Jorgensen, 2006). The Newton-Raphson method solves MLEs repeatedly by the standard least-squares methods (McCullagh & Nelder, 1989; Wedderburn, 1974). Both these methods result in parameter estimates that are alike. However, the estimated covariance matrix parameters may somewhat differ. "This is because the Fisher Scoring is based on the expected information matrix and New-Raphson is based on the observed information matrix" (SAS, 2019).

4.5 Logistic Model Selection and Checking

4.5.1 Model Selection

4.5.1.1 Wald Test

The test for the significance of each variable can be carried out once the model is fitted. The Wald test is the frequently used technique when a hypothesis test on a single parameter, β_j, is to be executed. The Wald Chi-square test statistic is :

$$z_0 = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \tag{4.36}$$

"This test statistic follows an approximate standard normal distribution. The standard error of $\hat{\beta}_j$ is the square root of the diagonal element of cov($\hat{\beta}$)" (MacKinnon, 2015). Numerous software packages take this Wald test statistic value and square it and does a comparison to a chi-square distribution with one degree of freedom (Archer et al., 2007; Bewick et al., 2005). Consequently, for bigger values of the test statistic, we reject the null hypothesis H_0: β_j =0 against the alternative that H_a : $\beta_j \neq 0$ and conclude that the variable that corresponds is significant to the model (Berndt & Savin, 1977). In other words, If the Wald Chi-square statistic is above the table value of chi-square, H_0 is not accepted, that suggests that the parameter is significantly different from zero.

4.5.1.2 ODDS RATIO

The ratio of the probability that a certain event of interest happens to the probability that it does not is called the odds. The odds can be explained in favor of an event of interest or against the event. The number of failures to the number of successes is the odds against an event of interest. In this study, we focus on odds in favor of the event of interest, which is "the ratio of the probability of the event happening to the probability of the event not happening" (Bland & Altman, 2000).

The odds ratio measures the relationship between the dependent and the independent variables. "For a continuous variable in logistic regression the odds ratio denotes how the odds vary with a one unit increase in that continuous variable with everything else held constant" (Wilber & Fu, 2010). When a categorical predictor has more than one category, the odds ratio is interpreted as the change in the odds of an event for each category in comparison to the odds of an event for the reference category. The interpretation of the odds ratio is with reference to this category at all times. Suppose x, is a categorical independent variable coded 0 and 1. The odds ratio is given as:

$$\mathrm{OR} = \frac{\mathrm{Odds}_1}{\mathrm{Odds}_2} = \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(0)}{1-\pi(0)}}$$

An odds ratio equivalent to one implies no association between the exposure and outcome, odds ratio bigger than 1 implies that the exposure is associated with greater odds of the outcome and odds ratio smaller than 1 means the exposure is associated with lesser odds of the outcome (Glas et al., 2003).

The odds ratio of one means that the event will happen, and zero suggesting the event won't happen. If we look at a dichotomous response variable which represents the event, the OR is explained as the ratio of odds for those with X=1 to X=0. The log of then odds ratio is given by

$$\log\left(\widehat{OR}\right) = log(OR(x=1, x=0)) = \widehat{\beta_1} \tag{4.37}$$

Therefore:

$$\mathrm{OR}=\exp\left(\widehat{\beta_1}\right) \tag{4.38}$$

The odds ratio shows how the odds of the event varies as X changes from zero to one. If there is a continuous variable called X then as X enlarges by one unit, then the odds of risk factor increase by $\exp\left(\widehat{\beta_1}\right)$.

4.5.1.3 Confidence Interval for the Odds Ratio

The confidence intervals (CI) gives a better view of the sampling variability for the estimates (Cadeddu et al., 2012). The confidence interval for the intercept and the slope are based on Wald tests (Altman et al., 2013). The $100\left(1-\frac{\alpha}{2}\right)$% CI for the intercept is defined by

$$\beta_0 \pm \mathrm{Z}_{1-\frac{\alpha}{2}}\mathrm{se}(\beta_0) \tag{4.39}$$

The $100\left(1-\frac{\alpha}{2}\right)$% CI for the slope is given by

$$\beta_{\mathrm{j}} \pm \mathrm{Z}_{1-\frac{\alpha}{2}}\mathrm{se}\left(\beta_{\mathrm{j}}\right) \tag{4.40}$$

These confidence intervals are on the logit scale and must be modified by exponentiation to get the corresponding $100\left(1-\frac{\alpha}{2}\right)$%. The CI for OR linked with β_{j} where $j = 1,2,3,\ldots,p$. is given by

$$\exp(\beta_1 \pm \mathrm{Z}_{1-\frac{\alpha}{2}}\mathrm{se}(\beta_1)). \tag{4.41}$$

4.5.2 Model Checking

The goodness of fit for a logistic regression model as defined by Hosmer-Lemeshow (1989) assesses the effectiveness of the model in describing the outcome variable (Archer & Lemeshow, 2006).

The fitted model's residual variation is expected to be small, displaying no systematic tendency and follows the model's variability(Hosmer et al., 1997). It can be measured using the Hosmer - Lemeshow tests, Pearson chi-square statistic and deviance statistic. "The Hosmer-Lemeshow goodness-of-fit (GOF) statistic ${\chi_{HL}}^2$ is found by computing the Pearson Chi-square statistic from the g × 2 table of observed and estimated expected frequencies”(Hosmer et al., 1997). Where g is the number of groups. The Hosmer-Lemeshow statistics is given by

$$\chi_{HL}{}^2 = \sum_{k=1}^{g} \frac{\left(O_k - n'_k \bar{\pi}_k\right)^2}{n'_k \bar{\pi}_k (1-\bar{\pi}_k)}. \tag{4.42}$$

n'_k is the total number frequency of subjects in the k^{th} group and the total frequency of event outcome in the k^{th} group is given by O_k while $\bar{\pi}_k$ is the average estimated probability of a certain event outcome in the k^{th} group. If the correct model is the logistic regression model, then the chi-square distribution may be used to estimate $\chi_{HL}{}^2$ with $g-2$ df (n=2 is the default in SAS model statement for the lack of fit option). The null hypothesis being tested is H_0: good fit model against the alternative H_a: model is not a good fit. The large value of Hosmer-Lemeshow statistic (p-value < 0.05) suggests a lack of fit of the model (Bewick et al., 2005; Fagerland & Hosmer, 2012).

4.5.2.1 Validating predicted probabilities

Understanding how much the predicted probabilities agree with the outcomes is imperative. We want to have a dependable model that will maximize the chance as well as the sensitivity of recognizing the individuals who require some intervention (Aldrich et al., 1984). The proportion of individuals that are not classified correctly as having outcome failure should have a reduction. A cut-off value that minimizes the misclassification probabilities of the individuals must be identified. Below is an example of how the classification is completed.

Table 4.1: Classification table

Correct classification

		y =1	y=0	Total
Predicted Classification	ŷ=1	a	b	a+b

	ŷ=0	c	d	c+d
		a+c	b+d	n

The probability of classifying individual with outcome failure incorrectly is known as False positive rate or 1− specificity which is calculated as

$$F_{pr} = \frac{b}{b+d}. \tag{4.43}$$

The probability of classifying individual with the outcome success incorrectly is known as False negative rate estimated as

$$F_{nr} = \frac{c}{a+c}. \tag{4.44}$$

$\hat{y}_i$ is the predicted response of the i^{th} individual and y_i is the response of the i^{th} individual. The probability of classifying an individual with the outcome of child death correctly is sensitivity given by

$$S_s = \frac{a}{a+c}. \tag{4.45}$$

The probability of classifying an individual with the outcome of child survival correctly is called Specificity and given by

$$S_p = \frac{d}{b+d}. \tag{4.46}$$

The Receiver Operating Characteristic curve may be utilized to present how accurate the prediction of the model is, graphically(Zou et al., 2007). "The Receiver Operating Characteristic curve is produced by plotting the sensitivity against the false positive rate at certain threshold settings. The Receiver Operating Characteristic curve compares two operating characteristics FPR and TPR as the criterion changes" (Metz, 2006).

The Receiver Operating Characteristic provides the measure of model ability to categorize between subjects that have experienced the outcome as opposed to the ones who didn't. This gives a plot of sensitivity versus 1-specificity (FPR) as displayed in Figure 4.1. A curve in the 45 degrees line, meaning area under the curve is 0.5, demonstrates that classification is at random

(Fawcett, 2006). The bigger the deviation of the curve is from the 45 degrees line to the left, the better is the model's accuracy to predict. This implies that the models prediction accuracy could be calculated by the total area under the Receiver Operating Characteristic curve (AUC) (Pearce & Ferrier, 2000).

A better diagnostic accuracy of the test is shown a bigger area under the curve. Suppose logistic models were fitted, and the first model produced D with AUC of 0.5, the second model produced C with AUC of 0.7 also the third model produced B with AUC of 0.9. C < B in Figure 4.1. suggests that the model B has the better predicting accuracy compared to the model C. We can classify the third model as the better model because it has a great diagnostic correctness and therefore has stronger accuracy. An AUC equal to 0.5 is bad as the test won't distinguish between falsely classified as positive outcomes and ones correctly classified positive outcomes.

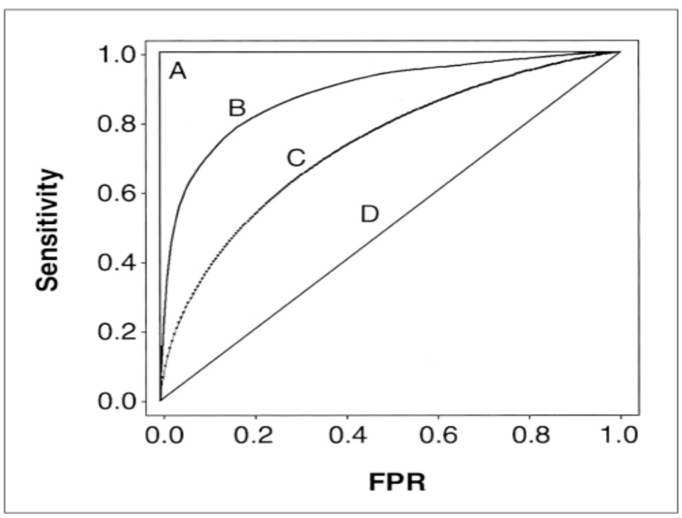

Figure 4.1: Sensitivity against False positive rate.

Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2698108/

Taylor and Krawchuk (2005) stated that the predicting accuracy of 0.6 to 0.7 suggests moderate prediction; and between 0.7 to 0.8 denoted acceptable prediction; and a range between 0.8 and 0.9 denotes excellent prediction. Range of 0.5-0.6 suggests bad prediction and anything <5 makes the test not useful (Taylor & Krawchuk, 2005).

4.6 Fitting the Logistic Regression Model

4.6.1 Fitting a LR model with a dichotomous response

The next part looks at the logistic regression, with a dichotomous outcome of child mortality status of children below age 5 in Lesotho. The Akaike information criterion for the reduced model containing only intercepts is bigger than the full model with the intercept and covariates; this means that the model fitted describes the data better (Table 4.2. and Table 4.3.).

Table 4.2: Logistic regression model fit statistics for binary response for LDHS 2009 data.

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	2571.559	1355.002
SC	2577.853	1568.991
-2 Log L	2569.559	1287.002

Table 4.3: Logistic regression model fit statistics for binary response for LDHS 2014 data.

Statistics for Model Fit		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	1611.01	788.288
SC	1617.11	994.033
-2 Log L	1609.063	720.288

Table 4.4. below shows the model evaluation for LR model for the LDHS 2009 data. The Likelihood Ratio Statistic tests the total significance of the LR model. The likelihood ratio statistic has a $p< 0.0001$and is 1282.5573. The score test is 1881.3265 with P-value < 0.0001, and the Wald test value is 637.0295 with a P-value < 0.0001 which supports the results found with the likelihood ratio test.

Table 4.4: Model evaluation for the LR Model LDHS 2009 data.

Model evaluation parameters	Chi Square	DF	Pr > ChiSq
Overall significance			

Likelihood Ratio	1282.5573	33	<.0001
Score	1881.3265	33	<.0001
Wald	637.0295	33	<.0001
Goodness of fit test			
Hosmer-Lemeshow Goodness-of-Fit	9.1035	8	0.3336
Association of Predicted Probabilities and Observed Responses			
Percent Concordant	92.8	Somers' D	0.856
Percent Discordant	7.2	Gamma	0.856
Percent Tied	0.00	Tau-a	0.152
Pairs	1417158	c	0.928

In Table 4.4, the complete fitted logistic model is significant as the p-value is less than 0.05. There is significant contribution of predictor variables in the prediction of the probability of child mortality for children under five. The H-L test for of this model is 9.1035 with a p-value equals to 0.3336, which indicates that the model is a good fit to the data since the p-value is above 5%. The model validation is also an important aspect to be checked. The extent to which the probabilities that are predicted correspond with the actual probabilities was shown by utilizing a classification table with a limit of 0.5 (Melesse et al., 2016).

To evaluate the predicted and observed probability values, the measures of association were used. The Goodman Krushkal Gamma statistic had a value of 0.856, the c-statistic had 0.928 and the Somer's D statistic had a value of 0.856. The values of the statistics were close to one, where 0 means no association and 1 means perfect association, which supports that there was a strong association between the predicted and observed probabilities.

The predictive accuracy of the model was evaluated through the ROC and the goodness-of-fit was tested with the Hosmer-Lemeshow test. Table 4.5. below shows the evaluation of the model for LR model for the LDHS 2014 data. The Likelihood Ratio test tests the altogether significance of the LR model. The likelihood ratio statistic is equals 888.775 with P-value less than 0.0001. The score test has a value of 1507.48 with a p< 0.0001 and the value of the Wald test is 393.191 with a P-value less than 0.0001 supports the results found with the likelihood ratio test.

Table 4.5: Model evaluation for the LR Model LDHS 2014 data.

Model evaluation parameters	Chi-Square	DF	PR > ChiSq
Overall significance			
Likelihood Ratio	888.775	33	<.0001
Score	1507.48	33	<.0001
Wald	393.191	33	<.0001
Goodness of fit test			
Hosmer-Lemeshow Goodness-of-Fit	7.1164	7	0.4169
Association of Predicted Probabilities and Observed Responses			
Percent Concordant	94.6	Somers' D	0.892
Percent Discordant	5.4	Gamma	0.892
Percent Tied	0	Tau-a	0.118
Pairs	650045	C	0.946

The overall logistic model fitted is significant as the p-value< 0.05. This indicates there's a significant contribution of independent variables in prediction of the probability of under-five child mortality. The H-L test of this model is 7.1164 with p-value equal to 0.4169, which proves that the model is a good fit to the data since the p-value is above 5%. The value of Goodman Krushkal Gamma statistic was 0.892, the c-statistic had a value of 0.946 and the Somer's D statistic had a value of 0.892. The values of these statistics were close to one, where 0 means no association and 1 means perfect association, which supports that there was a strong association between the observed and predicted probabilities.

Table 4.6. displays the Type 3 Analysis for the effects for the LDHS 2009 data. This table gives the hypothesis tests for each variable in the LR model with a certain degree of freedom test for the general effect of the categorical variables.

Table 4.6: Type 3 Analysis of Effects for the Logistic Regression for LDHS 2009 data.

Main Effect	DF	Wald Chi-Square	Pr > ChiSq
Current age of child	4	217.8648	<.0001
Mother's age	1	39.5	<.0001
Sex of child	1	3.0236	0.0821
Weight of the child at birth	2	4.1828	0.1235
Currently breastfeeding	1	0.0032	0.9547
Marital status	1	0.0666	0.7963
Wealth index	2	10.3145	0.0058
Education attainment	2	6.1451	0.0463
Mothers currently working	1	2.2678	0.1321
Type of place of residence	1	7.8409	0.0051
Region	9	8.393	0.4951
Birth order number	2	8.5894	0.0136
Number of children living	2	51.2765	<.0001
Significant interaction effect			
Age of child and Currently breastfeeding	4	27.5594	<.0001

The Wald chi-square test statistics and the p-values are displayed in table 4.6 above. The outcome shows that the overall effect of the categorical variables: wealth index, education attainment, type of place of residence, current age of child, birth order number and number of children living have significant effects on the under 5 child mortality. These results also indicate that the continuous variables: mother's age has significant effect statistically on the response variable with a $p<0.001$.

This also shows that categorical variables: currently breastfeeding, marital status, sex of child, mother's currently working status, weight of the child at birth and region do not have significant effect on the probability of child death using on the 2009 LDHS data used. The two-way interaction term of age of child and currently breastfeeding was found significant.

Table 4.7:Type 3 Analysis of Effects for the Logistic Regression for LDHS 2014 data.

Main Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Current age of child	4	92.5269	<.0001
Mother's age	1	11.2613	0.0008
Sex of child	1	0.7893	0.3743
Weight of the child at birth	2	6.866	0.0323
Currently breastfeeding	1	0.001	0.975
Marital status	1	2.0755	0.1497
Wealth index	2	2.9234	0.2318
Education attainment	2	6.9019	0.0317
Mothers currently working	1	0.4608	0.4972
Type of place of residence	1	4.6312	0.0314
Region	9	7.0953	0.6272
Birth order number	2	13.6976	0.0011
Number of children living	2	46.6576	<.0001
Significant interaction effect			
Age of child and Currently breastfeeding	4	45.2315	<.0001

The Wald chi-square test statistics and the p-values are displayed in the table 4.7 above. The outcome shows the overall effect of the categorical variables: education attainment, place of residence, weight of child, birth order number, current age of child and number of children living were found have significant effects on the under 5 child mortality. These results also indicate that the continuous variables: mother's age has significant effect on the response variable having a p-value less than 0.0008. This also shows that categorical variables: sex of child, currently breastfeeding, marital status, wealth index, mother's currently working, and region do not have significant effect on the probability of child death using the 2014 LDHS data. The two-way interaction term of age of child and currently breastfeeding was found significant.

4.6.2 Multiple Logistic Regression Model

In the paper, the probability of child mortality is modeled as a function of explanatory variables that are described in previous chapter. SAS PROC LOGISTIC is used to fit the model; The multiple logistic models were then fitted with all variables.

Table 4.8: Logistic regression model coefficients, standard errors, odds ratios, confidence intervals and significant two-way interaction effects for LDH 2009 data.

Main Effects	Estimate	SE	p-value	OR	95% CI	
					Lower	Upper
Intercept	-6.8255	0.8059	<.0001	0.001		
Current age of child (ref. 0)						
1 year	-1.7908	0.3637	<.0001	0.167	0.0818	0.340
2 years	-2.0566	0.7311	0.0049	0.128	0.0305	0.536
3 years	-2.4809	1.019	0.0149	0.084	0.0114	0.617
4 years	-13.9359	282.4	0.9606	0.000		
Mother's age	0.1039	0.0165	<.0001	1.109	1.074	1.146
Sex of child (ref. Male)						
Female	-0.2652	0.1525	0.0821	0.767	0.569	1.034
Weight of the child at birth (ref. Large)						
Average	-0.0595	0.2056	0.7723	0.942	0.630	1.410
Small	0.3332	0.2457	0.1751	1.395	0.862	2.259
Currently breastfeeding (ref. Yes)						
No	3.254	0.1975	<.0001	25.894	17.58	38.13
Marital status (ref. Not Married)						
Married	0.0477	0.1848	0.7963	1.049	0.730	1.507
Wealth index (ref. Rich)						
Middle	0.0274	0.2595	0.916	1.028	0.618	1.709
Poor	0.6176	0.2348	0.0085	1.854	1.170	2.938
Education attainment (ref.Higher Education)						
No Education	1.4505	0.6975	0.0376	4.265	1.087	16.737
Up to Secondary Education	1.2312	0.5052	0.0148	3.425	1.273	9.220
Mothers currently working (ref. Employed)						
Unemployed	-0.2481	0.1648	0.1321	0.780	0.565	1.078
Type of place of residence (ref.Urban)						
Rural	0.7545	0.2694	0.0051	2.127	1.254	3.606
Region (ref. Thaba-Tseka)						
Berea	0.2107	0.3636	0.5623	1.235	0.605	2.518
Butha-Bothe	0.4197	0.3632	0.2478	1.522	0.747	3.101
Leribe	0.3937	0.3282	0.2303	1.482	0.779	2.821
Mafeteng	0.8426	0.337	0.0124	2.322	1.200	4.496
Maseru	0.296	0.3442	0.3897	1.344	0.685	2.640
Mohale's Hoek	0.4083	0.3345	0.2223	1.504	0.781	2.898
Mokhotlong	0.1429	0.3277	0.6629	1.154	0.607	2.193
Qacha's-Nek	0.2235	0.3708	0.5467	1.250	0.605	2.586
Quthing	0.5136	0.3481	0.1401	1.671	0.845	3.306
Birth order number (ref. first birth)						
2-4 births	0.7851	0.2684	0.0034	2.193	1.296	3.710
>4 births	0.774	0.493	0.1165	2.168	0.825	5.699
Number of children living (ref. <2 children)						
2-4 children	-1.9223	0.2691	<.0001	0.146	0.086	0.248
>4 children	-2.2057	0.5056	<.0001	0.110	0.041	0.297
Significant Interactions Effects						

Current age of child (ref. 0 years) and currently breastfeeding (ref. yes)						
1 year’s vs Not Breastfeeding	-1.4091	0.4335	0.0012	0.244	0.1045	0.572
2 years vs Not Breastfeeding	-2.983	0.8107	0.0002	0.051	0.0103	0.248
3 years vs Not Breastfeeding	-3.0337	1.1067	0.0061	0.048	0.0055	0.421

Interpretation of the Coefficient of the Model and the Odds Ratio (2009 DATA)

Table 4.8. illustrates the parameter estimates, confidence intervals, odds ratios, standard errors and p-values. The variables that are significant had p-values that are less than 0.05.

The effect of a mother with no education was observed as positively associated with child mortality with a p-value equals 0.0376. The odds ratio that corresponds was 4.265 and the 95 percent CI(1.087 ; 16.737). The odds of dying for a child that comes from mother that has no education were 4.265 times the odds of dying for a child that comes from mother with higher education level. The effect of education of a mother with up to secondary education was also observed as positively associated with child mortality with a p-value equals 0.0145. The odds ratio that corresponds was 3.425 and the 95 percent CI(1.273 ; 9.220). The odds of dying for a child from mother with up to secondary education were 3.425 times the odds of dying for a child from mother with higher education level. The mother’s age was positively associated with child mortality with an estimate of 0.1039 and p-value less than 0.0001. The effect of not breastfeeding was also positively associated with child mortality with a p-value equals 0.0001 and the odds ratio that corresponds was 25.894 and the 95% CI (17.58 ; 38.13). The odds ratio being 1.109 and the 95% CI (0.569 ; 1.034) for mother’s age indicates that the odds of dying in under five children rises by 10.9% for a single unit increase in the mother’s age. The odds of dying for a child who was not breastfeed were 25.894 times the odds of dying for a child from a mother who breastfed.

The effect of the age of child 1 year was observed as negatively associated with child mortality with a p-value equals 0.0001. The odds ratio that corresponds is 0.167 and the 95 percent confidence interval (0.0818 ; 0.340). The odds of dying for a child who is 1 years old were 0.167 times the odds of dying for a child who was 0 years old. The effect of the age of child 2 year was observed as negatively associated with child mortality with a p-value equals 0.0049. The odds ratio that corresponds was 0.128 and the 95 percent confidence interval (0.0305 ; 0.536). The odds of dying for a child who is 2 years old were 0.128 times the odds of dying for a child who was 0 years old. The effect of the age of child 3 years was observed to be negatively associated with child mortality with a p-value equals 0.0149. The odds ratio that corresponds was 0.084 and the 95 percent confidence interval (0.0114 ; 0.617). The odds of dying for a child who is 3 years old were 0.084 times the odds of dying for a child that was 0 years old.

The effect of being poor was observed as positively associated with child mortality with a p-value equals 0.0085. The odds ratio that corresponds was 1.854 and the 95 percent confidence interval: 1.170 ; 2.938. The odds of dying for a child who comes from a poor household were 1.854 times the odds of dying for a child who comes from a rich household. The effect of rural area was observed as positively associated with child mortality with a p-value equals 0.0051. The odds ratio that corresponds was 2.127 and the 95% CI: 1.254 ; 3.606. The effect of childbirth order number two to four (2-4) was established to be negatively associated with child mortality with a p-value equals 0.0034. The odds ratio that corresponds was 2.193 and the 95 percent confidence interval: 1.296 ; 3.710. The odds of dying for a child with birth order between two and four were 2.193 times the odds of dying for a child with birth order number is less than two.

The effect of region (Mafeteng) was positively associated with child mortality with a p-value equals 0.0125. The odds ratio that corresponds was 2.322 with a 95 percent confidence interval: 1.200 ; 4.496. The odds of dying for a child who is from Mafeteng were 2.322 times the odds of dying for a child who is from Thaba-Tseka. The effect of the number of living children that is between two and four (2-4) was negatively associated with child mortality with a p-value equals 0.0001. The odds ratio that corresponds was 0.146 with a 95% confidence interval: 0.086 ; 0248. The odds of dying for a child from a mother with 2-4 children alive were 0.146 times the odds of dying for a child from a mother with less than two children alive. The effect of the number of living children that is above four was also negatively associated with child mortality with a p-value equals0.0001. The odds ratio that corresponds was 0.110 with a 95 percent confidence interval: 0.041 ;0.297. The odds of dying for a child from a mother who has more than four children alive were 0.532 times the odds of dying for a child from a mother with less than two children alive.

The significant two-way interaction effect was between the age of child and breastfeeding. The effect of age of child 1 year and not breastfeeding compared to a 0-year-old child that is breastfed was negatively associated with the child mortality with a p-value equals 0.0012. The odds ratio that corresponds was 0.244. The odds of child death to the effect of child aged 1 years and a child that is not breastfed is 0.244 times larger than the odds of child death to the effect of child aged 0 and were breastfed. The effect of age of child 2 year and not breastfeeding compared to a 0-year-old child that is breastfed was negatively associated with the child mortality with a p-value equals0.0002. The odds ratio that corresponds was 0.051. The odds of child death to the effect of child aged 2 years and a child that is not breastfed is 0.051 times larger than the odds of child death to the effect of child aged 0 and were breastfed. The effect of age of child 3 year and not breastfeeding compared to a 0-year-old child that is breastfed was negatively associated with the child mortality with a p-value equals 0.0061. The odds ratio that corresponds was 0.048. The odds of child death to the effect of child aged 3 years and a child that is not breastfed is 0.048 times larger than the odds of child death to the effect of child aged 0 and were breastfed.

Table 4.9: Logistic regression model coefficients, standard errors, odds ratios, confidence intervals and significant two-way interaction effects for LDHS 2014 data.

Main Effects	**Estimate**	**Standard Error**	**p-value**	**Odds Ratio**	**95% Confidence Limits**	
					Lower	Upper
Intercept	-7.9509	1.0328	<.0001			
Current age of child (ref. 0)						
1 year	-1.7704	0.6174	0.0041	0.1703	0.05077	0.5710
2 years	-0.1648	0.655	0.8013	0.8481	0.2349	3.062
3 years	-11.887	218.1	0.9565	6.879E-06	1.538E-191	3.076E+180
4 years	-12.3844	263.6	0.9625	4.183E-06	1.740E-230	1.005E+219
Mother's age	0.0763	0.0228	0.0008	1.079	1.032	1.129
Sex of child (ref. Male)						
Female	0.185	0.2082	0.3743	1.203	0.800	1.809
Weight of the child at birth (ref. Large)						
Average	0.2224	0.2847	0.4346	1.249	0.715	2.182
Small	0.8002	0.3359	0.0172	2.226	1.152	4.300
Currently breastfeeding (ref. Yes)						
No	4.0135	0.2815	<.0001	55.3402	31.8730	96.0857
Marital status (ref. Not Married)						
Married	0.3729	0.2589	0.1497	1.452	0.874	2.412
Wealth index (refresh)						
Middle	0.024	0.3081	0.9378	1.024	0.560	1.874
Poor	0.4463	0.305	0.1434	1.563	0.859	2.841
Education attainment (ref.Higher Education)						
No Education	1.5217	1.2334	0.2173	4.58	0.408	51.375
Up to Secondary Education	1.3877	0.5287	0.0087	4.006	1.421	11.29

Mothers currently working (ref. Employed)						
Unemployed	0.1726	0.2542	0.4972	1.188	0.722	1.956
Type of place of residence (ref. Urban)						
Rural	0.6367	0.2956	0.0314	1.889	1.058	3.372
Region (ref. Thaba-Tseka)						
Berea	0.1679	0.5029	0.7385	1.183	0.441	3.169
Butha-Bothe	-0.1034	0.506	0.8381	0.902	0.334	2.431
Leribe	0.7626	0.4815	0.1133	2.144	0.834	5.509
Mafeteng	0.7008	0.5106	0.1699	2.015	0.741	5.482
Maseru	0.1703	0.4862	0.7261	1.186	0.457	3.075
Mohale's Hoek	0.322	0.4799	0.5022	1.38	0.539	3.535
Mokhotlong	0.5305	0.484	0.2731	1.7	0.658	4.389
Qacha's-Nek	0.3529	0.5594	0.5281	1.423	0.475	4.26
Quthing	0.5872	0.4998	0.24	1.799	0.675	4.791
Birth order number (ref. first birth)						
2-4 births	1.2497	0.3586	0.0005	3.489	1.728	7.046
>4 births	1.9592	0.6384	0.0021	7.094	2.03	24.791
Number of children living (ref. <2 children)						
2-4 children	-2.4477	0.3647	<.0001	0.086	0.042	0.177
>4 children	-3.2406	0.6997	<.0001	0.039	0.01	0.154
Significant Interactions Effects						
Current age of child (ref. 0 years) and currently breastfeeding (ref. yes)						
1 year’s vs Not Breastfeeding	-1.7139	0.6847	0.0123	0.1802	0.0471	0.6894
2 years vs Not Breastfeeding	-5.2049	0.7963	<.0001	0.0055	0.0012	0.0261

Interpretation of the Coefficient of the Model and the Odds Ratio (2014 DATA)

Table 4.9. illustrates the parameter estimates, standard errors, odds ratios, p-values and confidence intervals for the 2014 data. The variables that are significant had p-values that were less than 0.05.

The effect of education of mother with up to secondary education was also observed as positively associated with child mortality with a p-value equals 0.0087. The odds ratio that corresponds was 4.006 with 95 percent CI (1.421 ; 11.29). The odds of dying for a child from a mother with up to secondary education were 4.006 times the odds of dying for a child from a mother with higher education level. The effect of not breastfeeding was positively associated with child mortality with a p-value equals 0.0001. The odds ratio that corresponds was 55.3402 with 95 percent confidence interval (31.873 ; 96.0857). The odds of dying for a child from a mother that does not breastfeed were 55.3402 times the odds of dying for a child from a mother that breastfeeds. The mother’s age was positively associated with child mortality (estimate of 0.0763). The odds ratio of 1.079 with 95% confidence interval(1.032 ; 1.129) for mother’s age indicates that the odds of dying in under five children rises by 7.9% for a single unit increase in the mother's age.

The effect of the age of child 1 year was observed as negatively associated with child mortality with a p-value equals 0.0041. The odds ratio that corresponds was 0.1703 with 95 percent confidence interval (0.05011 ; 0.57100). The odds of dying for a child who is 1 years old were 0.1703 times the odds of dying for a child that was 0 years old. The effect of a small weight child at birth was positively associated with the mortality with a p-value equals 0.0172. The odds ratio was 1.249 with 95 percent confidence interval(0.715 ; 2,182). The odds of dying for a child who is small in weight at birth is 1.249 times the odds of dying for a child that is large at birth. The effect of rural area was observed as positively associated with child mortality with a p-value equals 0.0314. The odds ratio that corresponds was 1.889 with 95 percent confidence interval(1.058 ; 3.372). The odds of dying for a child that is from a rural area is 1.889 times the odds of dying for a child that is from an urban area.

The effect of child-birth order number two to four (2-4) was observed as positively associated with child mortality with a p-value equals 0.0005. The odds ratio that corresponds was 3.489 with 95 percent confidence interval(1.728 ; 7.046). The odds of dying for a child whose birth order number between two and four (2-4) were 3.489 times the odds of dying for a child that has birth order number less than two. The effect of child-birth order number >4 was observed as positively associated with child mortality with a p-value equals 0.0021. The odds ratio that corresponds was 7.092 with 95 percent CI (2.03 ; 24.791). The odds of dying for a child whose birth order number was >4 were 7.092 times the odds of dying for a child whose birth order number is less than two.

The effect of the number of living children that is between two and four (2-4) was observed as negatively associated with child mortality with a p-value equals 0.0001. The odds ratio that corresponds was 0.086 with 95 percent confidence interval(0.042 ; 0.177). The odds of dying for a child from a mother with 2-4 children alive were 0.086 times the odds of dying for a child from a mother with less than two children alive. The effect of the number of living children that is above four was also negatively associated with child mortality with a p-value equals 0.0001. The odds ratio that corresponds was 0.039 with 95 percent confidence interval(0.01 ;0.154). The odds of dying for a child from a mother with more than four children alive were 0.039 times the odds of dying for a child from a mother with less than two children alive.

The significant two-way interaction effect was observed between age of child and breastfeeding. The effect of age of child 1 year and not breastfeeding compared to a 0-year-old child that is breastfed was negatively associated with the child mortality with a p-value equals 0.0123. The odds ratio that corresponds was 0.1802. The odds of child death for a child aged 1 year and is not breastfed is 0.1802 times that of a child aged 0 and is breastfed. The effect of age of child 2 year and not breastfeeding compared to a 0-year-old child that is breastfed was negatively related to the child mortality with a p-value equals 0.0001. The odds ratio that corresponds was 0.0055. The odds of child death for a child aged 2 years and that is not breastfed is 0.0055 times that of a child aged 0 and is breastfed.

4.6.3 Prediction Accuracy of the Logit Model

The prediction accuracy validates the logit model. Checking how much the predicted probability agrees with outcomes is crucial. Figure 4.2 illustrates the ROC curve of the fitted model using the 2009 LDHS data. The proportion of the probabilities predicted correctly is the area under the ROC curve and for this model 92.79% (c= 0.9279) of the probabilities are correctly predicted. This is a very good predictive accuracy in this model (Centor & Schwartz, 1985).

For a good predictive model, the area under the curve must be maximum (almost 1). The area under the curve (AUC) is known as the concordance, c, in SAS. To obtain better predicting power c should be above 0.5 (Austin & Steyerberg, 2012; Pearce & Ferrier, 2000).

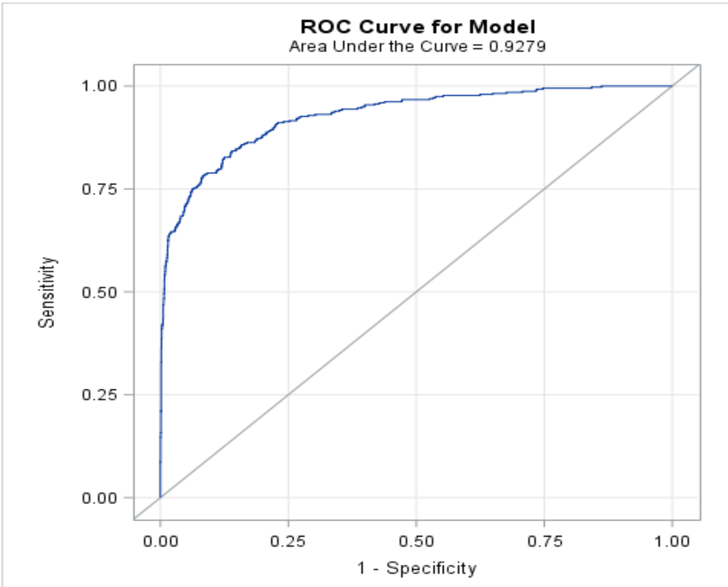

Figure 4.2: (ROC) Receiver Operating Characteristic curve for logit model.

Figure 4.3. shows the ROC curve of the fitted model using the 2014 LDHS data. The area under the curve is the proportion of the probabilities correctly predicted and for this model 94.61% (c=0.9461) of the probabilities are predicted correctly. This is a good predictive accuracy in this model.

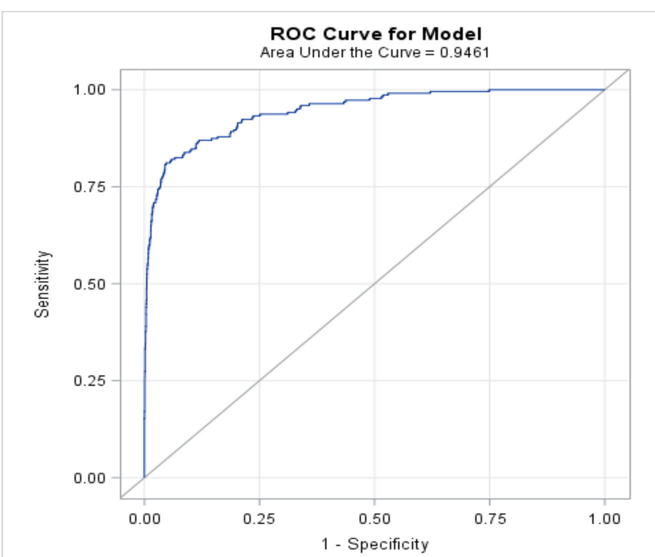

Figure 4.3: (ROC)Receiver Operating Characteristic curve for logit model.

4.6.4 Logistic Regression Diagnostic Plots

The Standardized Pearson residuals, Deviance residuals, and influence measures that assist in understanding how observations behave in the model. The influence diagnostics which were generated by using influence option in procedure PROC LOGISTIC while fitting a logistic regression model to the data are shown in Figure 4.4. and 4.5. The value of the diagnostic is represented by the vertical axis on each plot, and the case number of the observation is represented by the horizontal axis. Observations that are further away from zero are influential observations and are poorly accounted for by the model

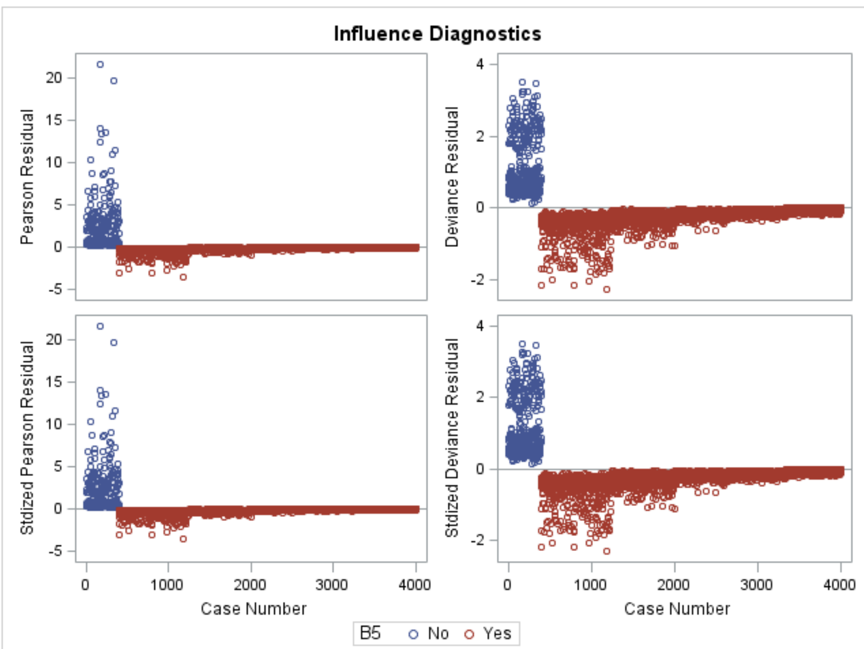

Figure 4.4: Influence diagnostic from the logistic regression model (2009).

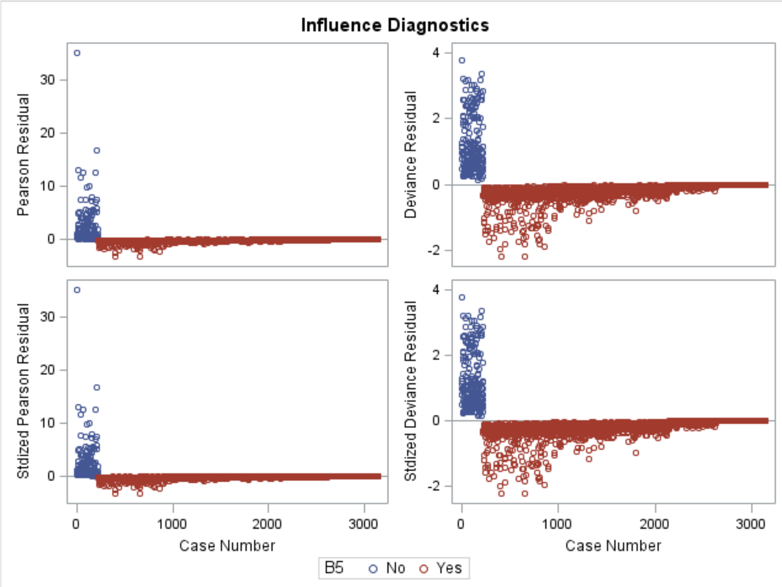

Figure 4.5: Influence diagnostic from the logistic regression model (2014)

4.7 Survey Logistic Regression Model

Logistic regression is very beneficial in modelling data that has an outcome that is dichotomous however, it is not fitting for modelling data acquired by a complex survey which accounts for clustering, stratification, and weights. "Logistic regression models are applied to examine data from the complex sampling designs are called the survey logistic regression models in this study" (Rabe-Hesketh & Skrondal, 2006; Yirga, 2018).

The theory in survey logistic regression models and ordinary LR models is equivalent. Data being collected using SRS to fit the model is the assumption made in OLR and complexity of the survey design is considered in the SLR (Ayele et al., 2012). If the level of complexity of the design is disregarded when modeling, then the standard errors could be overestimated or underestimated which would result to narrower or wider confidence intervals. The results may not be significant but will seem to be statistically significant and this can result in us getting biased estimates (Hilbe, 2009). Simple random sampling designs make the assumption that the probability of inclusion in the sample is equivalent for all units in a population (Burger & Silima, 2006). The importance in this is that survey logistic regression accounts for clustered correlated observations whereas simple logistic regression does not (Bieler et al., 2010). "The survey logistic regression model and the ordinary logistic regression model are identical when the data is from simple random sample. The calculation of standard errors are affected by clustering and stratification whilst the sampling weight will influence the calculation of the point estimate" (Bennett et al., 1991).

Survey logistic regression models the relationship between binary dependent variables and the set of explanatory variables by using the sampling design information (Roberts et al., 1987). Including the effects of sampling design in the analysis of the data leads to accurate estimation of the standard errors and variabilities (Skinner & Wakefield, 2017). The advantages of sample surveys are stated below:

- They are cost effective, but the cost depends on the survey mode.
- Capability of collecting data from many respondents
- Faster and therefore time saving (developed in less time than other methods).
- Production of quality and accurate population estimates and are feasible (Kish, 1965; Wyse, 2012).

The one important advantage of stratification is that the parameters can be estimated for each stratum and the survey is easier to administer. The variance of the estimator of a population total is reduced by diving the population into strata (Ding et al., 1996; Lumley, 2004).

The section that follows will explain the methods of parameter estimation for survey logistic.

4.7.1 Estimation of Parameters

The independent assumption does not hold in complex survey design, when clusters are being drawn, they could introduce correlation among observations. There needs to be appropriate estimation of the standard errors associated with the model coefficients. In short, the complexity of the sampling design must be considered in the analysis . The standard error constructed by the assumption of a simple random sample will most likely underestimate the real value (Thomas & Heck, 2001; Winship & Radbill, 1994).

In the present situation the primary sample units (PSUs) were sampled in each stratum in the first stage. The household was sampled in the second stage. $y_{hijk}(h = 1,2,\ldots,H_{kji}\,;\ i = 1,2,\ldots,n_{kj};\ j = 1,2,\ldots,m_k\,;\ k = 1,2,\ldots,K)$ is the response variable that is equal to one if the event occurred in h^{th} individual within i^{th}household within j^{th} primary sample unit nested within k^{th} stratum, and zero otherwise. The total number of observations is stated as:

$$n = \sum_{k=1}^{K} \sum_{j=1}^{m_k} n_{kj} \tag{4.46}$$

The probability is given by $\pi_{kjih} = P(y_{hijk} = 1)$ if the event occurred in h^{th} individual within i^{th} household within j^{th} primary sample unit nested within k^{th} stratum and the probability is given by $1 - \pi_{kjih} = P(y_{hijk} = 0)$ if the event did not occur in h^{th} individual within i^{th}household within j^{th} primary sample unit nested within k^{th} stratum. Thus, the log-likelihood function in this case is given by

$$l(\boldsymbol{\beta}; y) = \sum_{k=1}^{K} \sum_{j=1}^{m_k} \sum_{i=1}^{n_{kj}} \sum_{h=1}^{H_{kji}} \left\{ y_{kjih} \log\left(\frac{\pi_{kjih}}{1-\pi_{kjih}}\right) - \log\left(\frac{1}{1-\pi_{kjih}}\right) \right\} \tag{4.47}$$

and the survey logistic regression model is written as:

$$logit(\pi_{kjih}) = \log\left(\frac{\pi_{kjih}}{1-\pi_{kjih}}\right) = \boldsymbol{X'}_{kjih}\beta \tag{4.48}$$

with $h = 1,2,\ldots,H_{kji}\,;\ i = 1,2,\ldots,n_{kj};\ j = 1,2,\ldots,m_k\,;\ k = 1,2,\ldots,K$. Where X_{kjih} is the vector corresponding to the characteristics of h^{th} individual within i^{th}household within j^{th} primary sample unit nested within k^{th} stratum and β is a vector of unknown parameters of the model.

If all design variables are included in the model as independent variables, the inference about the effects of the factors in the fitted model will be reliable (Moineddin et al., 2007; Pfeffermann, 1993). The Pseudo maximum likelihood technique named the weighted maximum likelihood which considers the sampling design and various sampling weights in estimating β are used to estimate the parameters β of the logistic regression model . (Asparouhov & Muthen, 2006; Hosmer & Lemeshow, 2000). The pseudo maximum likelihood function is used to obtain the parameter estimates. "The pseudo maximum likelihood is computed as the product of individual contributions to the likelihood" (Geys1 et al., 1997).

"The Pseudo maximum likelihood function for the contribution of one observation in complex sampling design" (Archer et al., 2007) is given by:

$$\pi_{kjih}{}^{W_{kjih}y_{kjih}} \left(1-\pi_{ijh}\right)^{\left(1-W_{kjih}y_{kjih}\right)} \tag{4.49}$$

Therefore, the pseudo maximum likelihood function with weights is given by:

$$L(\boldsymbol{\beta};Y) = \prod_{k=1}^{K} \prod_{j=1}^{m_k} \prod_{h=1}^{n_{kj}} \prod_{h=1}^{H_{kji}} \pi_{kjih}{}^{W_{kjih}y_{kjih}} \left(1-\pi_{kjih}\right)^{\left(1-W_{kjih}y_{kjih}\right)} \tag{4.50}$$

The log-likelihood function is given by:

$$l(\boldsymbol{\beta};Y) = \sum_{k=1}^{K} \sum_{j=1}^{m_k} \sum_{i=1}^{n_{kj}} \sum_{h=1}^{H_{kji}} \left\{ W_{kjih} y_{kjih} \log\left(\frac{\pi_{kjih}}{1-\pi_{kjih}}\right) - \log\left(\frac{1}{1-\pi_{kjih}}\right) \right\} \tag{4.51}$$

The maximum likelihood estimates, $\hat{\beta}$, is attained by making the score equation above equal 0 and solve for β using the iterative methods of Newton Raphson and Fisher scoring (Institute, 2015).

4.7.2 Survey Logistic Model Selection and Checking

4.7.2.1 Model selection

Logistic regression utilizes forward selection, backward elimination and stepwise selection procedures to select the variables that best fit the data set. In the survey logistic procedure in SAS PROC SURVEYLOGISTIC method for analyzing complex survey data, the variable selection procedures such as backward selection, forward selection, score and stepwise selection are not yet available. One must manually add or remove one variable in the model at a time by using the type 3 analysis of effects and observe the effect of the variables left. These analyses are often used when the effect of one predictor variable is influenced by the effect of another predictor variable. One can remove variable that is not significant and refit the model without that variable. This manual approach can be done until all remaining variables in the model are significant. Thereafter, one may consider including interaction terms amongst the variables in the model.

4.7.2.2 Model Fit Test

The goodness-of-fit of the model is tested using the AIC and the SC criteria. These measures may be used to compare two nested models when determining the better model that explains the data set. These criteria are used because SAS PROC SURVEYLOGISTIC does not give the Hosmer-Lemeshow statistics and plots (Lumley & Scott, 2014). The smaller the BIC and AIC of the full model in comparison to the corresponding BIC and AIC of the model that is reduced, then the better the full model. More information of the BIC and AIC criteria for model selection can be found in section 4.3 in this study.

4.7.2.3 Predictive Accuracy of the Model

The PROC SURVEYLOGISTIC is used to fit binary response models to data, and it generates statistics on the models' ability to predict, like Sommer's D, c, Kendall's Tau-a and Goodman-Kruskal Gamma. All these statistics are between 0 and 1. The larger value corresponds to a strong association between observed and predicted values. These measures of association are as discussed before.

4.7.3 Design Effects

The clustering, stratification, and weighting of selected cases affects the sampling variance of a survey statistic. The accuracy of the variance estimate can be increased by the stratification but the precision can be decreased by weighting and clustering (Lumley, 2004).

Stratified sampling is a technique that incorporates the splitting of a population into smaller sub-groups known as strata (Yadav et al., 2019). A sample is independently taken from each stratum. Stratified random sampling is when a SRS is taken in each stratum. (Levy & Lemeshow, 2013).

Stratified sampling is used:

- To decrease sampling error in comparison to the simple random sampling for greater precision
- For its convenience in administration
- When different sampling procedure are needed for different parts of population
- When separate estimates are needed at strata level.

There are two principal advantages when comparing simple random sampling and stratified sampling. It may need more administrative effort than a simple random sample and the analysis has more computationally complexity. A more informative discussion about sampling methods can be discovered in many literatures (Cochran, 1977, 2007; Hibberts et al., 2012).

The precision of the parameter estimates is determined by the sample size and sampling design. "The problem in complex sample design is that sampling errors for survey estimates cannot be simply worked out using the formulae obtained in statistical books" (Kish & Frankel, 1974). Design effect, DEFF, is the loss of effectiveness in the use of complex instead of simple random sampling. The ratio of actual variance divided the variance expected with SRS defines the design effect (Levy & Lemeshow, 2013; Shackman, 2001). The design effect is a technique that is extensively used in survey sampling for planning a sample design in estimation and analysis (Rust & Rao, 1996; Särndal et al., 2003). The DEFF option can be used in the model statement. PROC SURVEYLOGISTIC calculates the design effect for the regression coefficients. The design effect is given by:

$$\mathrm{DEFF} = \frac{\text{variance under the complex design}}{\text{variance under simple random sampling}}. \tag{4.52}$$

The design effect is used to find how much bigger the sample size or confidence interval should be. DEFF will usually range from one to three. However, it's possible for the design effect to be much higher.

One can also use DEFT given by the square root of DEFF. The DEFT, which is more preferred, may be used to decrease variability since DEFF is more variable than DEFT (Salganik, 2006). The DEFT illustrates by how much the sample confidence intervals (and SE) increase. DEFT equal to one denotes there's no effect of sample design on the standard error. DEFT that is bigger than one implies that the sample design increases the estimates standard error. DEFT that is less than 1 implies sampling design does not increase the estimates standard error.

4.8 Results of the Survey Logistic Regression Model

The analysis in this portion of the study applies the survey logistic regression model, that accounts for the complex survey design. Multiple logistic regression was fitted for the Lesotho DHS data using SAS. The model selected by the PROC LOGISTIC was fitted again using the PROC SURVEYLOGISTIC to see if estimates would differ when the complexity of the survey design is considered.

4.8.1 Model Checking and Prediction Accuracy of the Model

The PROC SURVEYLOGISTIC in SAS does not give the Hosmer-Lemeshow statistics and plots, so we use the AIC and SC to check if the model is a good fit. The AIC of the full model is smaller in comparison to the AIC of the model that is reduced; this means the model fitted is a good fit (Table 4.10. and Table 4.11.).

Table 4.10: Statistics for Model fit for the survey logistic regression(LDHS 2009).

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	2478.746	1296.561
SC	2484.970	1508.203

-2 Log L	2476.746	1228.561

Table 4.11: Statistics for Model fit for the survey logistic regression (LDHS 2014).

Model Fit Statistics		
Criterion	**Intercept Only**	**Intercept and Covariates**
AIC	1570.324	772.824
SC	1576.367	978.283
-2 Log L	1568.324	704.824

The tables below (Table 4.12. & 4.13.) shows that the p-values that correspond to the likelihood ratio, Wald tests and score tests are less than 0.05 (p<.0001) for the 2009 and 2014 LDHS data.

Table 4.12: Model evaluation for Survey logistic regression (LDHS 2009).

Model Evaluation parameters	F Value	Num DF	Den DF	Pr > F
Likelihood Ratio	41.22	27.439	10400	<.0001
Score	25.38	33	347	<.0001
Wald	397.51	33	347	<.0001
Association of predicted probabilities and observed response				
Percent Concordant	92.3	Somers' D	0.850	
Percent Discordant	7.3	Gamma	0.854	
Percent Tied	0.4	Tau-a	0.151	
Pairs	1417158	c	0.925	

This means that the fitted models are significant. There is a significant contribution of independent variables in the prediction of the children under 5 years dying in both periods. There is a high (92.5% and 94.4%) association between the predicted probabilities and the observed responses. The concordant rate is 92.3% in 2009 and 94.2% in the 2014 period as shown in both tables; these values tell us how good the model was in separating 0's and 1's. The Gamma statistic has a value of 0.854 and 0.893 indicates a positive association between variables and the Somer's D statistic is 0.850 in 2009 and 0.889 in 2014.

Table 4.13: Model evaluation for Survey logistic regression (LDHS 2014).

Model Evaluation parameters	F Value	Num DF	Den DF	Pr > F
Likelihood Ratio	26.82	23.9923	10930	<.0001
Score	186.68	33	345	<.0001

Wald	314.80	33	345	<.0001
Association of Predicted Probabilities and Observed Responses				
Percent Concordant	94.2	Somers' D	0.889	
Percent Discordant	5.3	Gamma	0.893	
Percent Tied	0.5	Tau-a	0.117	
Pairs	650045	c	0.944	

Table 4.14. illustrates that the two-way interaction term has a significant interaction effect on the response variable with p-value<.0001 in the 2009 data. The P-values shown in the table indicate that the continuous variable mother's current age was found to have a significant effect on the response. The overall effect of categorical variables: sex of the child, breastfeeding status, wealth index, current age of child, education attainment, type of resident and number of children alive have a significant effect on the probability of death of children.

Table 4.14: Type 3 analysis of effects for the LR with complex survey design (2009).

Main Effect	F Value	Num DF	Den DF	Pr > F
Current age of child	167.20	4	376	<.0001
Mother's age	38.88	1	379	<.0001
Sex of child	7.08	1	379	0.0081
Weight of the child at birth	0.36	2	378	0.6953
Currently breastfeeding	81.82	1	379	<.0001
Marital status	0.08	1	379	0.783
Wealth index	3.85	2	378	0.0221
Education attainment	6.5	2	378	0.0017
Mothers currently working	2.87	1	379	0.091
Type of place of residence	9.02	1	379	0.0029
Region	0.75	9	371	0.6665
Number of children living	24.21	2	378	<.0001
Birth order number	1.56	2	378	0.2125
Significant interaction effect				
Age of child and Currently breastfeeding	26.75	4	376	<.0001

Table 4.15. illustrates that the two-way interaction term has significant interaction effect on the response variable with a p<.0001 in the 2014 data. The P-values shown in the table shows that the continuous variable mother's current age was found to be significant on the response. The overall effect of categorical variables: current age of child, breastfeeding status, education

attainment, number of children living, and birth order number were found to have a significant effect on the probability of death of children.

Table 4.15: Type 3 analysis of effects for the LR with complex survey design (2014).

Main Effect	F Value	Num DF	Den DF	Pr > F
Current age of child	1112.59	4	374	<.0001
Mother's age	13.92	1	377	0.0002
Sex of child	1.53	1	377	0.2174
Weight of the child at birth	2.74	1	376	0.0660
Currently breastfeeding	48.43	1	377	<.0001
Marital status	2.65	1	377	0.1043
Wealth index	0.93	2	376	0.3963
Education attainment	4.03	2	376	0.0186
Mothers currently working	0.24	2	377	0.6233
Type of place of residence	1.13	1	377	0.2895
Region	0.91	1	369	0.5159
Number of children living	17.30	9	376	<.0001
Birth order number	5.64	2	376	0.0037
Significant interaction effect				
Age of child and Currently breastfeeding	58.01	4	374	<.0001

4.8.2 Interpretation of the Coefficients of the Model and Odds ratio .

Table 4.16: Survey logistic regression model estimation for main effects and significant interaction effects(2009).

Main Effects	**Estimate**	**SE**	**P-value**	**OR**	**CI**	
					lower	upper
Intercept	-7.4997	0.8683	<.0001	2.383		
Current age of child (ref. 0)						
1 year	-1.798	0.3856	<.0001	0.16563	0.07779	0.35267
2 years	-1.7655	0.7993	0.0278	0.171101	0.03572	0.81965
3 years	-2.8769	1.0551	0.0067	0.056309	0.00712	0.44535
4 years	-13.822	0.249	<.0001	9.94E-07	0.00000	0.00000
Mother's age	0.1166	0.0187	<.0001	1.12367	1.08323	1.16562
Sex of child (ref. Male)						
Female	-0.5237	0.1969	0.0081	0.592325	0.40268	0.87129
Weight of the child at birth (ref. Large)						
Average	0.0535	0.2372	0.8215	1.054957	0.66271	1.67936

Small	0.2233	0.2929	0.4462	1.250196	0.70414	2.21973
Currently breastfeeding (ref. Yes)						
No	-3.2276	0.1963	<.0001	0.039653	0.02699	0.05826
Marital status (ref. Not Married)						
Married	-0.0525	0.1904	0.783	0.948854	0.65332	1.37807
Wealth index (ref. Rich)						
Middle	-0.2383	0.2659	0.3706	0.787966	0.46792	1.32692
Poor	0.4223	0.2584	0.1031	1.525466	0.91928	2.53138
Education attainment (ref.Higher Education)						
No Education	1.8553	0.7255	0.0109	6.393616	1.54237	26.50358
Up to Secondary Education	1.8502	0.5132	0.0004	6.361092	2.32641	17.39307
Mothers currently working (ref. unemployed)						
Employed	0.2884	0.1702	0.091	1.334291	0.95581	1.86263
Type of place of residence (ref. Urban)						
Rural	1.122	0.3737	0.0029	3.07099	1.47631	6.38820
Region (ref. Thaba Tseka)						
Berea	0.1334	0.324	0.6808	1.142707	0.60554	2.15640
Butha-Bothe	0.3585	0.3425	0.2959	1.431181	0.73140	2.80051
Leribe	0.4511	0.347	0.1944	1.570038	0.79531	3.09944
Mafeteng	0.6986	0.3235	0.0314	2.010935	1.06667	3.79111
Maseru	0.3644	0.4187	0.3847	1.43965	0.63365	3.27086
Mohale's Hoek	0.3588	0.3337	0.2829	1.43161	0.74434	2.75344
Mokhotlong	0.2458	0.3276	0.4535	1.278644	0.67281	2.43001
Qacha's-Nek	0.2413	0.3543	0.4963	1.272903	0.63564	2.54907
Quthing	0.5531	0.4035	0.1712	1.738634	0.78839	3.83420
Number of children living (ref. <2 children)						
2-4 children	-2.1722	0.312	<.0001	0.113927	0.06181	0.20999
>4 children	-2.3386	0.561	<.0001	0.096463	0.03212	0.28966
Birth order number (ref. first birth)						
2-4 births	0.6146	0.348	0.0782	1.848917	0.93475	3.65713
>4 births	0.6532	0.5701	0.2526	1.92168	0.62864	5.87435
Significant Interactions Effects						
Current age of child (ref. 0 years) and currently breastfeeding (ref. Yes)						
1 year's vs No Breastfeeding	-1.5596	0.4395	0.0004	0.21022	0.08883	0.49749
2 years vs No Breastfeeding	-3.0271	0.8762	0.0006	0.048456	0.00870	0.26989
3 years vs No Breastfeeding	-3.0263	1.1614	0.0095	0.048495	0.00498	0.47239
4 years vs Breastfeeding	6.9118	0.8145	<.0001	0.000254	3.44181	5.33474

Table 4.16. shows the estimated coefficients, standard errors, and p-value for the logistic regression model. The output from SAS PROC SURVEYLOGISTIC for the LDHS data 2009 is given in the table; it displays that the effect of not breastfeeding and the child aged 1,2 and 3 was negatively associated with child mortality with a p-value <0.05. The odds ratios that correspond were 0.2102, 0.0485 and 0.0485. This means that the odds of child dying for a child aged 1 year

and that is not breastfed is 0.2102 times that of a child who is aged 0 that was breastfed. This implies that the odds of child mortality for a child aged 2 and3 years and a child that is not breastfed is 0.0485 times that of a child aged 0 that was breastfed. The effect of four years old child and not breastfeeding compared to a 0-year-old child that is breastfed was positively associated with the child mortality ($p<0.0001$). This implies that the odds of child mortality of for child aged 4 years and that is not breastfed is 0.000254 times that of the odds death for a child aged 0 that was breastfed. The results from Table 4.16. also show that the probability of child mortality has significant association with mother's current age, current age of child, breastfeeding, education attainment, type of residence , Mafeteng region and number of children living when we looked at the logistic regression model with complex sampling design.

The odds ratio of 1.12367 for mother's age shows that for a one-unit increase in mother's age, the odds of death in under five children increases by 12.37%. The mother's age was positively associated with child mortality. The effect of breastfeeding was established as negatively associated with the child mortality with a p-value<0.0001. The odds ratio that corresponds was 0.0397. The odds of dying for a child from a mother who is not breastfeeding were 0.0397 times the odds of dying for a child from a mother who does breastfeed. The effect of no education and up to secondary education was observed as positively associated with the child mortality ($p=0.0109$ and $p=0.0004$). The corresponding odds ratio was 6.39 and 6.36. The odds of dying for a child from a mother who has no education were 6.39 times the odds of dying for a child from a mother who has higher education whilst the odds of dying for a child from a mother who has up to secondary education were 6.36 times the odds of dying for a child from a mother who has higher education.

The effect of a child from rural area was positively associated with the child mortality ($p=0.0029$). The corresponding odds ratio was 3.08. The odds of dying for a child from a rural area were 3.08 times the odds of dying for a child from an urban area. The effect of a child from Mafeteng region was observed as positively associated with the child mortality ($p=0.0314$). The corresponding odds ratio was 2.01. The odds of dying for a child from Mafeteng region were 2.01 times the odds of dying for a child from Thaba Tseka region. The effect of the mother with number of children alive which is between two and four was negatively associated with child mortality with a p-value equals 0.0001. The odds ratio that corresponds was 0.114. The odds of dying for a child from a mother with 2-4 children alive were 0.114 times the odds of dying for a child from a mother with less than two children alive. The effect of the mother with number of children alive which is more than four was negatively associated with child mortality with a p-value equal 0.0001. The corresponding odds ratio was 0.096. The odds of dying for a child from a mother with more than four children alive were 0.096 times the odds of dying for a child from a mother with less than two children alive.

Table 4.17: Survey logistic regression model estimation for main effects and significant interaction effects(2014).

Main Effects	Estimate	SE	P-value	OR	CI	
					lower	upper
Intercept	-8.23	0.9991	<.0001	0.0003		
Current age of child (ref. 0)						
1 year	-1.651	0.6575	0.0125	0.1919	0.05288	0.69607
2 years	0.142	0.8744	0.871	1.1526	0.20767	6.39697
3 years	-11.3961	0.3361	<.0001	0	0.00001	0.00002
4 years	-11.8585	0.3498	<.0001	0	0.00000	0.00001
Mother's age	0.0838	0.0225	0.0002	1.0874	1.04050	1.13644

Sex of child (ref. Male)						
Female	0.3065	0.2481	0.2174	1.3587	0.83546	2.20952
Weight of the child at birth (ref. Large)						
Average	0.0617	0.3556	0.8622	1.0636	0.52979	2.13545
Small	0.6498	0.3955	0.1012	1.9152	0.88216	4.15777
Currently breastfeeding (ref. Yes)						
No	4.0252	0.3382	<.0001	55.9915	28.85629	108.64351
Marital status (ref. Not Married)						
Married	0.3992	0.2451	0.1043	1.4906	0.92201	2.40993
Wealth index (ref. Rich)						
Middle	0.0519	0.3198	0.8712	1.0533	0.56276	1.97133
Poor	0.3773	0.3408	0.269	1.4583	0.74776	2.84416
Education attainment (ref.Higher Education)						
No Education	2.0555	1.0172	0.044	7.8107	1.06374	57.35225
Up to Secondary Education	1.6252	0.5826	0.0055	5.0794	1.62142	15.91236
Mothers currently working (ref. employed)						
Unemployed	0.1305	0.2654	0.6233	1.1394	0.67727	1.91685
Type of place of residence (ref. Urban)						
Rural	0.3885	0.3663	0.2895	1.4748	0.71932	3.02360
Region (ref. Thaba-Tseka)						
Berea	0.2394	0.5073	0.6373	1.2705	0.47005	3.43394
Butha-Bothe	-0.1427	0.4836	0.7681	0.867	0.33603	2.23705
Leribe	0.822	0.4834	0.0899	2.275	0.88209	5.86771
Mafeteng	0.6337	0.4741	0.1822	1.8846	0.74413	4.77281
Maseru	0.0852	0.4633	0.8542	1.0889	0.43917	2.70004
Mohale's Hoek	0.2805	0.4009	0.4845	1.3238	0.60335	2.90451
Mokhotlong	0.4107	0.4316	0.342	1.5079	0.64711	3.51358
Qacha's-Nek	0.3921	0.4773	0.412	1.4801	0.58077	3.77201
Quthing	0.7713	0.4392	0.0799	2.1626	0.91436	5.11477
Number of children living (ref. <2 children)						
2-4 children	-2.6614	0.4786	<.0001	0.0699	0.02734	0.17847
>4 children	-3.7136	0.6894	<.0001	0.0244	0.00632	0.09420
Birth order number (ref. first birth)						
2-4 births	1.5878	0.5517	0.0042	4.893	1.65942	14.42744
>4 births	2.3406	0.6935	0.0008	10.3875	2.66803	40.44164
Significant Interactions Effects						
Current age of child (ref. 0 years) and currently breastfeeding (ref. Yes)						
1-year vs No Breastfeeding	-1.7114	0.7633	0.0255	0.1806	0.04046	0.80627
2 years vs No Breastfeeding	-5.591	1.052	<.0001	0.0037	0.00047	0.02933
3 years vs No Breastfeeding	4.9659	1.1275	<.0001	3.4376	15.73677	7.40561
4 years vs No Breastfeeding	-5.7933	0.4689	<.0001	0.003	0.00122	0.00764

Table 4.17. shows the output from SAS PROC SURVEYLOGISTIC for the LDHS data 2014; it illustrates that the effect of not breastfeeding and the child aged 1,2 and 4 was negatively associated with child mortality ($p < 0.05$). The odds ratios that correspond were 0.1806, 0.0037

and 0.0030. This implies that the odds of child death for a child aged 1 years that is not breastfed is 0.1806 times that of the odds of a child aged 0 that was breastfed. This also means that the odds of child mortality to the effect of child aged 2 and 4 years and a child that is not breastfed is 0.0037 and 0.0030 times larger than the odds of child death to the effect of child aged 0 that was breastfed. The effect of age of child 3 years and not breastfeeding compared to a 0-year-old child that is breastfed was positively associated with the child mortality ($p<0.0001$). This implies that the odds of child mortality for a child aged 4 years that is not breastfed is 143.44 times that the odds of for a child aged 0 that was breastfed.

The results from Table 4.17. also show that the probability of child mortality is significantly associated with mother's current age, current age of child(1 year,3 years and 4 years), breastfeeding, education attainment, birth order number and number of children living when we look at the survey logistic regression model. The odds ratio of 1.087 for mother's age means that for a single unit increase in mother's age, the odds of dying in children under five increases by 8.74%. The mother's age was observed as positively associated with child mortality. The effect of breastfeeding was observed as positively associated with the child mortality ($p<0.0001$). The corresponding odds ratio was 55.99. The odds of dying for a child from a mother who is not breastfeeding were 55.99 times the odds of dying for a child from a mother who does breastfeed.

The effect of no education and up to secondary education was positively associated with the child mortality ($p=0.044$ and $p=0.0055$). The corresponding odds ratios were 7.81 and 5.079. The odds of dying for a child from a mother who has no education were 7.81 times the odds of dying for a child from a mother who has higher education whilst the odds of dying for a child from a mother with up to secondary education were 5.079 times the odds of dying for a child from a mother with higher education. The effect of the number of children alive which is between 2 and four was observed to be negatively associated with child mortality with a p-value equals 0.0001. The odds ratio that corresponds was 0.0699. The odds of dying for a child from a mother with 2-4 children alive were 0.0699 times the odds of dying for a child from a mother with less than two children alive. The effect of the number of children alive which is more than four was observed as negatively associated with child mortality ($p =0.0001$). The corresponding odds ratio was

0.0244. The odds of dying for a child from a mother with more than four children alive were 0.0244 times the odds of dying for a child from a mother with less than two children alive.

The effect of a child with birth order number is between 2 and four was observed to be positively associated with child mortality with a p-value equals 0.0042. The corresponding odds ratio was 4.8930. The odds of dying of a child whose birth order number is between 2-4 were 4.8930 times the odds of dying for a child with birth order number is less than two. The effect of a child with birth order number is greater than four was observed to be positively associated with child mortality (p =0.0008). The corresponding odds ratio was 10.39. The odds of death of a child with birth order number is greater than 4 were 10.39 times the odds of death for a child with birth order number is less than two.

4.8.3 Comparing the Logistic and Survey Logistic Regression

The tables below compare the standard errors obtained from PROC SURVEYLOGISTIC procedure and PROC LOGISTIC procedure based on DEFF and DEFT. Since the sample was not from the SRS, the parameter estimates for both models are not the same but closer to each other. Tables (4.18. & 4.19.) shows the results for each significant effect in the research. The logistic regression assumes that the observations are independent, but for complex design this assumption is breached and so a better model may be fitted using PROC SURVEYLOGISTIC since the complexity of the design is accounted for.

Table 4.18. shows, the results for the 2009 data set. The effect of mother's age has the DEFF equals 1.2844 and DEFT equals 1.1333. This indicates that the confidence interval and standard error are 1.1333 times as larger as they would be for SRS.

Table 4.18: Design effects estimated for response variable (2009).

Significant Effects	Estimates(CSD)	SE	var(CSD)	P-value	SE	var(SRS)	DEFF	DEFT
Intercept	-7.4997	0.8683	0.7539	<.0001	0.8059	0.6495	1.1609	1.0774
Current age of child (ref. 0)								
1 year	-1.798	0.3856	0.1487	<.0001	0.3637	0.1323	1.1241	1.0602
2 years	-1.7655	0.7993	0.6389	0.0278	0.7311	0.5345	1.1953	1.0933
3 years	-2.8769	1.0551	1.1132	0.0067	1.019	1.0384	1.0721	1.0354

Mother's age	0.1166	0.0187	0.0003	<.0001	0.0165	0.0003	1.2844	1.1333
Sex of child (ref. Male)								
Female	-0.5237	0.1969	0.0388	0.0081	0.1525	0.0233	1.6671	1.2911
Currently breastfeeding (ref. No)								
Yes	-3.2276	0.1963	0.0385	<.0001	0.1975	0.0390	0.9879	0.9939
Education attainment (ref.Higher Education)								
No Education	1.8553	0.7255	0.5264	0.0109	0.6975	0.4865	1.0819	1.0401
Up to Secondary Education	1.8502	0.5132	0.2634	0.0004	0.5052	0.2552	1.0319	1.0158
Mothers currently working (ref. unemployed)								
Employed	0.2884	0.1702	0.0290	0.091	0.1648	0.0272	1.0666	1.0328
Type of place of residence (ref. Urban)								
Rural	1.122	0.3737	0.1397	0.0029	0.2694	0.0726	1.9242	1.3872
Region (ref. Thaba-Tseka)								
Mafeteng	0.6986	0.3235	0.1047	0.0314	0.337	0.1136	0.9215	0.9599
Number of children living (ref. <2 children)								
2-4 children	-2.1722	0.312	0.0973	<.0001	0.2691	0.0724	1.3443	1.1594
>4 children	-2.3386	0.561	0.3147	<.0001	0.5056	0.2556	1.2312	1.1096
Significant Interactions Effects								
Current age of child (ref. 0 years) and currently breastfeeding (ref. Yes)								
1 year's vs No Breastfeeding	-1.5596	0.4395	0.1932	0.0004	0.4335	0.1879	1.0279	1.0138
2 years vs No Breastfeeding	-3.0271	0.8762	0.7677	0.0006	0.8107	0.6572	1.1681	1.0808
3 years vs No Breastfeeding	-3.0263	1.1614	1.3488	0.0095	1.1067	1.2248	1.1013	1.0494

The effect of breastfeeding has the DEFF equals 0.9879 and DEFT equals 0.9939. The confidence interval and standard error are 0.9939 times smaller as they would be for SRS. The effect of living in Mafeteng has the DEFF equals 0.9215 and DEFT equals 0.9599. This indicates that the confidence interval and standard error are 0.9599 times that of the standard error for SRS.

The effect of the number of children alive that is more than four is negatively associated with child mortality and has the DEFF =1.2312 and DEFT =1.1096. The confidence interval and standard error must be 1.1096 times as large as they would be for SRS. The effects for the number of children alive which is between two and four is also negatively associated with the child mortality and has the DEFF=1.3443 and DEFT=1.1594. The confidence interval and standard error must be 1.1594 times as large as they would be for SRS.

The effect of no education is positively associated with child mortality and has the DEFF =1.0819 and DEFT =1.0401. The confidence interval and standard error must be 1.0401 times as large as they would be for SRS. The effects of up to secondary education is also positively associated with the child mortality and has the DEFF =1.0319 and DEFT=1.0158. The confidence interval and standard error must be 1.0158 times as large as they would be for SRS. The effect of age of child for 1 year old has the DEFF=1.1241 and DEFT=1.0602. The confidence interval and standard error interval must be 1.0602 times as large as they would be for SRS. The effect of age of child for 2 years old has the DEFF=1.1953 and DEFT=1.0933. The confidence interval and standard error must be 1.0933 times as large as they would be for SRS. The effect of age of child for 3 years old has the DEFF equals1.0721 and DEFT equals 1.0354. The confidence interval and standard error must be 1.0354 times as large as they would be for SRS. The effect of being employed has the DEFF equals 1.0666 and DEFT equals 1.0328. The confidence interval and standard error are 1.0328 times as large as they would be for SRS.

The effect of living in a rural area has the DEFF equals 1.9242 and DEFT equals 1.3872. The confidence interval and standard error are 1.3872 times as large as they would be for SRS . The interaction effect of age of child for 1 year old and the child is breastfed has the DEFF=1.0279 and DEFT=1.0138. The confidence interval and standard error must be 1.0138 times as large as they would be for SRS. The interaction effect of age of child (for 2 years old) and the child is breastfed has the DEFF=1.1681 and DEFT=1.0808. The confidence interval and standard error must be 1.0808 times as large as they would be for SRS.

The effect of age of child for 3 years old depends on whether the child is breastfed with DEFF=1.1013 and DEFT=1.0494. The confidence interval and standard error must be 1.0494 times as large as they would be for SRS (see Table 4.18).

In Table 4.19 , the results shown are for the 2014 data set. The effect of mother's age has the DEFF equals 0.9739 and DEFT equals 0.9868. The confidence interval and standard error are 0.9868 times as smaller as they would be for SRS.

Table 4.19:Design effects estimated for response variable (2014).

Significant Effects	Estimate(CSD)	SE	var(CSD)	P-value	SE	var(SRS)	DEFF	DEFT
Intercept	-8.23	0.9991	0.9982	<.0001	1.0328	1.0667	0.9358	0.9674
Current age of child (ref. 0)								
1 year	-1.651	0.6575	0.4323	0.0125	0.6174	0.3812	1.1341	1.0649
Mother's age	0.0838	0.0225	0.0005	0.0002	0.0228	0.0005	0.9739	0.9868
Currently breastfeeding (ref. Yes)								

No	4.0252	0.3382	0.1144	<.0001	0.2815	0.0792	1.4434	1.2014
Education attainment (ref. Higher Education)								
No Education	2.0555	1.0172	1.0347	0.044	1.2334	1.5213	0.6802	0.8247
Up to Secondary Education	1.6252	0.5826	0.3394	0.0055	0.5287	0.2795	1.2143	1.1019
Number of children living (ref. <2 children)								
2-4 children	-2.6614	0.4786	0.2291	<.0001	0.3647	0.133	1.7222	1.3123
>4 children	-3.7136	0.6894	0.4753	<.0001	0.6997	0.4896	0.9708	0.9853
Birth order number (ref. first birth)								
2-4 births	1.5878	0.5517	0.3044	0.0042	0.3586	0.1286	2.3669	1.5385
>4 births	2.3406	0.6935	0.4809	0.0008	0.6384	0.4076	1.1801	1.0863
Significant Interactions Effects								
Current age of child (ref. 0 years) and currently breastfeeding (ref. Yes)								
1-year vs No Breastfeeding	-1.7114	0.7633	0.5826	0.0255	0.6847	0.4688	1.2428	1.1148
2 years vs No Breastfeeding	-5.591	1.052	1.1067	<.0001	0.7963	0.6341	1.7453	1.3211

The effect of breastfeeding has the DEFF equals 1.4434 and DEFT equals 1.2014. The confidence interval and standard error are 1.2014 times as large as they would be for SRS. The effect of no education is positively associated with child mortality has the DEFF=0.6802 and DEFT=0.8247. The confidence interval and standard error must be 0.8247 times as small as they would be for SRS. The effects of up to secondary education is also positively associated with the child mortality and has the DEFF=1.2143 and DEFT=1.1019. The confidence interval and standard error must be 1.1019 times as large as they would be for SRS.

The effect the number of children alive that is greater than four is negatively associated with child mortality has the DEFF=0.9708 and DEFT=0.9853. The confidence interval and standard error must be 0.9853 times as small as they would be for SRS. The effects for the number of children alive which is between two and four is also negatively associated with the child mortality and has the DEFF=1.7222 and DEFT=1.3123. The confidence interval and standard error must be 1.3123 times as large as they would be for SRS. The effect of age of child for a 1-year-old has the DEFF=1.1341 and DEFT=1.0649. The confidence interval and standard error must be 1.0649 times as large as they would be for SRS.

The effect of childbirth order number above four which is positively associated with the child mortality has the DEFF=1.1801 and DEFT=1.0863. The confidence interval and standard error are 1.0863 times large as they would be for SRS. The effect of childbirth order number between 2 and 4 is positively associated with the child mortality and has the DEFF=2.3669 and DEFT=1.5385. The confidence interval and standard error are 1.5385 times large as they would be for SRS. The effect of age of child for 1 year old depends on whether the child is breastfed with DEFF=1.2428 and DEFT=1.1148. The confidence interval and standard error must be 1.1148 times as large as they would be for SRS. The effect of age of child for 2 years old depends on whether the child is breastfed with DEFF equals1.7453 and DEFT equals 1.3211. The confidence interval and standard error must be 1.3211 times as large as they would be for SRS.

From the results above, we notice that most of the design effects values are higher than one. This means that there was a variance under-estimation when the logistic regression was used, that presumes data was sampled using SRS. This confirms that the under the survey logistic, the CI and standard errors are larger. Hence, using the model like survey logistic regression is good since it considers the features of survey design.

4.9 Shortcomings of the SURVEYLOGISTIC Procedure

The SLR accounts for how complex the survey designs are, but it may have limitations due to unobtainability of Hosmer-Lemeshow goodness of fit test in the SURVEYLOGISTIC procedure. The variable selection procedures are not available thus one is required to select variable manually which can consume time when many variables are involved, and possible errors may result while picking variables. The SURVEYLOGISTIC procedure does not have the ‘output’ option statement which enables more analysis of data, like testing the link function’s appropriateness, detecting outliers and influence. The model is forced to be chosen by using of the AIC and the SC criteria, both of which present a penalty to the -2log-likelihood of having a lot of parameters.

The results obtained through Survey Logistic Regression modelling tends to be unbiased, since it considers the complex design of the sample in the analysis. However, variability due to correlation amongst the elements selected from the same cluster also needs to be considered. The next chapter introduces the Generalized linear mixed model (GLMM), an expansion of the GLM that fits outcomes with distributions that are not normal and includes the random effects additionally to the fixed effects to be analyzed.

Chapter 5

5 Generalized Linear Mixed Models

5.1 Introduction

Chapter 4 made use of survey logistic regression modelling under generalized linear models to investigate the factors associated with under five child mortality. This chapter provides us with an alternative method for modelling under five child mortality, given that our data was collected from a survey that incorporated stratification and cluster sampling, that could lead to variability and correlation amongst subjects from households within the same cluster.

"Generalized linear mixed models also abbreviated as GLMMs are an extension of generalized linear models to allow response variables from different distributions" (McCulloch & Neuhaus, 2005), such as binary response in this study. This extension allows inclusion of both fixed and random effects as well as generalized linear models (Bruin, 2006; Møller & Waagepetersen, 2007). The generalized linear mixed model focuses more on the inverse link function rather than the link function to model the relationship between the conditional mean and linear predictor and also includes nonlinear mixed models (Kachman, 2000).

The GLMM is an important model in solving the over-dispersion problems and makes inference of the population heterogeneity. This chapter will focus on using the GLMM to investigate the risk factors related to death of children below the age 5. We will discuss the structure of the GLMM, provide various methods of estimation for both the random effects and fixed effects parameters and thereafter, apply the model to our data.

5.2 Generalized Linear Mixed Model

The classification of GLMMs includes various important types of statistical models. This contains:

• The linear models have an identity link function, no random effects and normal distribution

• The generalized linear models have no random effects

• The linear mixed models have a normal distribution, random effects and identity link function (Schabenberger, 2005)

GLMMs are given by the general form:

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{Zu} + \boldsymbol{\epsilon} \tag{5.1}$$

$$\mathbf{u} \sim \mathrm{N}(0, \mathrm{D}) \tag{5.2}$$

$$\boldsymbol{\epsilon} \sim \mathrm{N}(0, \mathrm{R}) \tag{5.3}$$

where:

$N \times 1$ response vector of responses is given by **Y**

$N \times p$ design matrix of fixed effect parameter given by **X**.

$\boldsymbol{\beta}$ is $p \times 1$ vector of fixed effects that are not known.

$N \times q$ design matrix for random effects given by **Z**.

U is $q \times 1$ vector of unknown random effects.

$\boldsymbol{\epsilon}$ is $N \times 1$ vector of error terms for responses.

D is $q \times q$ variance covariance component of random effects.

R is $N \times N$ matrix for the residuals.

We define random effects as variance and covariance of the observation and fixed effects as the expected value of the observation (Wolfinger, 1993). It may be assumed that the observations on the same unit are correlated. The Linear mixed models address the problem of covariation between measures on the same unit (Littell et al., 2000). U and ϵ are assumed independent and the covariance structure for the response vector is obtained from:

$$\boldsymbol{V}(\boldsymbol{Y}) = \boldsymbol{ZGZ}' + \boldsymbol{R} \tag{5.4}$$

Equation (5.4) gives the structure of V(Y) as a function of G and R. Where $\boldsymbol{ZGZ'}$ signifies the between-patient section of the covariance structure, and R denotes the within-patient portion (Littell et al., 2000).

The random effects are usually not estimated but predicted and the variance components are estimated. The diagonal elements of matrix G is the variance component for each random effect while off-diagonal elements are covariances that exist between different dimensions (Meyer, 1989). If there is K random effects in the model, then G will have K × K elements that is the variance components for K random effects.

5.3 Model Formulation

Given the random effects parameter u_i, the response variable, y_{ij} are presumed to be independent and have a distribution like the exponential family (Gueorguieva, 2001). The random effect parameter u_i is drawn independently and has a distribution of f(u_i, D).

$$f_{y_{ij}/u_i}(y_{ij}|u_i) = exp\left(\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{a(\phi)} + c(y_{ij}, \phi)\right) \tag{5.5}$$

the scale parameter is ϕ.

The conditional mean $\mu_{ij} = E(y_{ij}|u_i)$ is modeled with both random and fixed effects as:

$$g(\mu_{ij}) = \eta_{ij} = X'_{ij}\beta + Z'_{ij}u_i \tag{5.6}$$

The link function is g(.), u_i is a vector of random effects and the linear predictor is given by η(.).

5.4 Parameter Estimation

The maximum likelihood method is the preferred estimation method for the GLMMs. "Nevertheless, due to the nonlinearity of the model and the existence of random effects, finding the likelihood of the model needs a challenging integration with respect to the random effects' distribution" (Lee et al., 2018; McCulloch, 1997).

In a GLMM model, the likelihood is defined by:

$$L(\beta, G, \phi) = \prod_{i=1}^{N} f_i\left(Y_{ij}/\beta, G, \phi\right)$$

$$= \prod_{i=1}^{N} \int f_i\left(Y_{ij}/\beta, G, \phi\right) . f(U_i, G) d\, u_i \tag{5.7}$$

where, $f_i\left(Y_{ij}/\beta, G, \phi\right) = \int \ \prod_{j=1}^{n_i} f_i\left(Y_{ij}/\beta, G, \phi\right) . f(U_i, G) du_i$ (McCulloch & Neuhaus, 2005).

In the GLMMs, to get the MLE of the parameters, we must maximize the log-likelihood function with respect to u_i and β. To get the maximum likelihood estimates, we must integrate over the u_i random effects. However, in general, the integral has no closed form, resulting in an intractable maximum likelihood estimation problem (Liu, 2015). The numerical integration of the likelihood for the GLMM is intractable and results in inconsistent estimates and is computationally infeasible for high dimensions of the random effects (Nadeem, 2013).

Many other approximation methods have been suggested to deal with this integration problem to obtain the maximum likelihood estimate. These approximate maximum likelihood methods are best categorized as estimating equation techniques, and they include penalized quasi-likelihood, marginal quasi-likelihood, Gaussian quadrature, Laplace approximation (McCulloch, 1997) and various similar techniques (Sun & Ronnegard, 2011). Methods based on different approaches are discussed in the following sections.

5.4.1 Penalized Quasi-Likelihood

Difficulties in computing the maximum likelihood in GLMMs has resulted in simpler approximations to the likelihood function. The penalized quasi-likelihood is one of those approximations (PQL) (Breslow, 2004). The PQL method functions are fine when the data are approximately normally distributed but can be biased for data that highly deviates from a normal distribution. The method is more flexible than the full maximum likelihood procedure as only the first two moments of the conditional density need to be specified in terms of the GLMM model parameters(Breslow & Clayton, 1993).

The PQL method approximates the integral of the quasi-likelihood by decomposing data into the mean and the error terms using the Taylor series expansion of the mean (Wolfinger & Lin, 1997). Consider the following:

$$Y_{ij} = \mu_{ij} + \epsilon_{ij}$$

$$= h\left(x'_{ij}\beta + z'_{ij}u_i\right) + \epsilon_{ij} \tag{5.8}$$

where, $\mathrm{h}\left(x'_{ij}\beta + z'_{ij}u_i\right) = \mathrm{g}^{-1}\left(x'_{ij}\beta + z'_{ij}u_i\right)$ is the inverse of the link function. The error terms follow an appropriate distribution with mean zero and variance equal to $\mathrm{Var}\left(\mathrm{Y}_{ij}\right) = \phi\,\mathrm{v}\left(\mu_{ij}\right)$. To get an approximation of the mean, and then the parameters, the Taylor series expansion of Equation (5.8) is implemented. Assuming the natural link function, $\mathrm{v}\left(\mu_{ij}\right) = \mathrm{h}'\left(x'_{ij}\beta + z'_{ij}u_i\right)$ where h' is the derivative with respect to μ_{ij}. When the Taylor expansion is carried out with respect to estimates $\hat{\beta}$ and $\hat{u_i}$, the method is known to as Penalized Quasi-Likelihood (Goldstein & Rasbash, 1996). Then this yields the following result.

$$\begin{aligned} Y_{ij} \approx\ & h\left(x'_{ij}\beta + z'_{ij}\hat{u}_i\right) \\ & + \mathrm{h}'\left(x'_{ij}\beta + z'_{ij}\hat{u}_i\right) x'_{ij}\left(\beta - \hat{\beta}\right) \\ & + \mathrm{h}'\left(x'_{ij}\beta + z'_{ij}\hat{u}_i\right) z'_{ij}\left(u_i - \hat{u}\right) + \epsilon_{ij} \end{aligned}$$

$$=\widehat{\mu_{ij}} + \mathrm{v}\left(\widehat{\mu_{ij}}\right) x'_{ij} \left(\beta - \hat{\beta}\right) + \mathrm{v}\left(\widehat{\mu_{ij}}\right) z'_{ij} \left(u_i - \hat{u}\right) + \epsilon_{ij} \qquad (5.9)$$

where μ_{ij} is equal to its current predictor h $\left(x'_{ij}\beta + z'_{ij}\hat{u}_i\right)$ for the conditional mean $\mathrm{E}\left(\mathrm{Y}_{ij}|u_i\right)$.
In vector form:

$$\boldsymbol{Y_i} = \widehat{\boldsymbol{\mu_i}} + \widehat{\boldsymbol{V_i}}\, \boldsymbol{X_i}\left(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}\right) + \widehat{\boldsymbol{V_i}}\, \boldsymbol{Z_i}(\boldsymbol{u_i} - \widehat{\boldsymbol{u}}) + \boldsymbol{\epsilon_{ij}} \qquad (5.10)$$

where $\boldsymbol{X_i}$ and $\boldsymbol{Z_i}$ are appropriate design matrices and $\widehat{\boldsymbol{V_i}}$ is the diagonal matrix with elements , $\mathrm{v}\left(\widehat{\mu_{ij}}\right) = \mathrm{h}'\left(x'_{ij}\hat{\beta} + z'_{ij}\widehat{u_i}\right)$. Re-ordering the above expression and multiplying by $\widehat{\boldsymbol{V}}_{\boldsymbol{i}}^{\boldsymbol{-1}}$ gives

$$\boldsymbol{y_i}^* = \widehat{\boldsymbol{V}}_{\boldsymbol{i}}^{\boldsymbol{-1}}\left(\boldsymbol{y_i} - \widehat{\boldsymbol{\mu_i}}\right) + \boldsymbol{X_i}\,\widehat{\boldsymbol{\beta}} + \boldsymbol{Z_i}\,\widehat{\boldsymbol{u}}$$

$$\approx \boldsymbol{X_i}\,\boldsymbol{\beta} + \boldsymbol{Z_i}\,\boldsymbol{u} + \boldsymbol{\epsilon_i}^* \qquad (5.11)$$

For $\boldsymbol{\epsilon_i}^* = \widehat{\boldsymbol{V}}_{\boldsymbol{i}}^{\boldsymbol{-1}}\boldsymbol{\epsilon_i}$ and has a mean of zero. This can be seen as an LMM for a pseudo response $\boldsymbol{y_i}^*$ with error term $\boldsymbol{\epsilon_i}^*$. This will allow the use of algorithm for fitting generalized linear mixed models.

<u>The Algorithm</u>

The iteration algorithm for model fitting is as follows:

Step 1: Set the values for β,G and ϕ. Calculate empirical Bayes estimator for u_i and pseudo-response y_i^*.

Step 2: Based on pseudo-response y_i^* fit the model and update β ,G and ϕ accordingly.

These steps above are replicated until we achieve convergence, and the estimates that result are called penalized quasi-likelihood (Handayani et al., 2017).

5.4.2 Marginal Quasi-Likelihood

Marginal quasi-likelihood (MQL) is less challenging computationally in comparison to MLE and predominantly applies to longitudinal data (Rabe-Hesketh et al., 2002). This technique is like the PQL method although it's regularly used when interest is on the marginal relationship between the covariables and the outcome of interest (Breslow & Clayton, 1993). It is based on a linear Taylor expansion of the mean around current estimate $\hat{\beta}$ for fixed effects but around $\widehat{u_i} = 0$ for random effects (Masangwi et al., 2010). This gives an equivalent expansion as for PQL, but in this case the current predictor of $\widehat{\mu_{ij}}$ will be in form $h\left(x'_{ij}\beta\right)$.
The form of the pseudo-data is now:

$$\boldsymbol{y_i}^* = \widehat{\boldsymbol{V}}_{\boldsymbol{i}}^{-\mathbf{1}}\left(\boldsymbol{y_i} - \widehat{\boldsymbol{\mu_i}}\right) + \boldsymbol{X_i}\,\widehat{\boldsymbol{\beta}} \tag{5.12}$$

This satisfies the approximate LMM:

$$\approx \boldsymbol{X_i}\,\boldsymbol{\beta} + \boldsymbol{Z_i}\,\boldsymbol{u} + \boldsymbol{\epsilon_i}^* \tag{5.13}$$

The fitting of the model is done as in PQL, by repetition between fitting an approximate LMM and the calculation of the pseudo data for these pseudo data (Wolfinger & O'connell, 1993); However, the resulting estimates are known as marginal quasi-likelihood estimates.

5.4.2.1 Discussion of Marginal Quasi-Likelihood and Penalized Quasi-Likelihood

There isn't a lot of difference between marginal quasi-likelihood and penalized quasi-likelihood. Both of these techniques in the linear predictor don't include the random effects (u_i) and produce improved estimates with higher-order Taylor expansion (Breslow & Clayton, 1993). They have similar properties and are based on similar ideas. The MQL totally disregards the random effects variation in linearizing the mean and only gives good results if the variance of the random effects is very small. The MQL will give biased estimates when the number of measurements increases while PQL will give consistent values with increased measurements. PQL and MQL gives results that are very bad for binary outcomes with a few iterated measurements per cluster (Codd, 2014; Molenberghs & Verbeke, 2006).

5.4.3 Laplace Approximation

Laplace approximation is the usual technique used when the exact likelihood function is hard to assess (Kuk, 1999). This method is based on an approximation of the integrand and is one of the options once the exact likelihood function is challenging to calculate (Shun & McCullagh, 1995). The integrands are approximated, and the objective is to find traceable integrals so that closed form expressions can be acquired. This makes it possible to numerically maximize the approximated likelihood (Onar, 2014).

Suppose we want to find the integral of the form:

$$I = \int e^{-q(x)} dx \tag{5.14}$$

here a known and unimodal function is given by $q(x)$, and a $q \times 1$ vector of variables is given by x and second-order Taylor expansion of $q(x)$ about $\hat{b}$ is obtained by:

$$q(x) \approx q(\hat{x}) + \frac{1}{2}(x - \hat{x})' q''(\hat{x})(x - \hat{x}) \tag{5.15}$$

where $q''(\hat{x})$ is the Hessian of q calculated at $(\hat{x})$ and the first-order term of the Taylor expansion vanishes since the expansion is done with respect to $\hat{x}$ and $q''(\hat{x})$=$-l''(\beta, G, \phi)|_{x-\hat{x}}$ is a Hessian of the log-likelihood assessed at $\hat{x}$. By using the approximation in the Laplace approximation, the quadratic term changes to:

$$I = (2\pi)^{\frac{q}{2}}\, q''(\hat{x})|^{-\frac{1}{2}e^{q(\hat{x})}} \tag{5.16}$$

and the marginal log-likelihood changes to:

$$l(\hat{x}, y) = log \int exp\Big(l(\theta, \hat{x}, y) - \frac{1}{2}(x - \hat{x})' \, q''(\hat{x}) \, (x - \hat{x}) \, dx\Big)$$

$$= l(\theta, \hat{x}, y) - \frac{1}{2} log \left| \frac{q''(\hat{x})}{2\pi} \right| \tag{5.17}$$

If we have many repeated measures per subject, this results in good approximation.
The Laplace method is amongst the computational techniques used for estimation in GLMMs. It is the fastest when computing, but there hasn't been a clear analysis if it has enough accuracy.

5.4.4 Gaussian Quadrature

As explained above the Laplace approximation depends on a linearization technique of the integrand and another different approach to this is the approximation of the integral or numerical integration. The Gaussian quadrature is a numerical method of approximating a hard to solve integral of the marginal likelihood (Tuerlinckx et al., 2006). (Liu & Pierce, 1994) explains the Gaussian quadrature as follows:
Given an integral of the form:

$$\int_{-\infty}^{\infty} f(x)\Phi(x)dx \tag{5.18}$$

where $\Phi(x)$ is the density of the multivariate normal distribution. The Gaussian quadrature approximation is:

$$\int_{-\infty}^{\infty} f(x)exp(-x^2)dx \tag{5.19}$$

which is approximately

$$\sum_{i=1}^{q} w_i f(x_i) \tag{5.20}$$

where q is the order of the approximation, the nodes x_i are solutions to the qth order polynomial and w_i are well chosen weights. (Molenberghs & Verbeke, 2005) says that the approximation will be more accurate if we have higher q.

5.5 Generalized Linear Mixed Models (GLIMMIX)

The models can be fitted using statistical software programs like SAS, GENSTAT, and many more. In the current research, we focus on SAS applications since this is the software that will be used for the analysis of our data. The Statistical Analysis Software procedure PROC GLIMMIX accommodates features of GLMMs. This procedure combines the two procedures namely PROC GENMOD and PROC MIXED (Vonesh, 2012).

The GLIMMIX procedure considers the random effects and permits for population-average and subject-specific inference (Bolker et al., 2009). The GENMOD procedure only allows marginal inference (Neal & Simons, 2007). "The response can have a non-normal distribution, but the MIXED procedure assumes that the response is normally distributed" (Schabenberger, 2005). The SAS PROC GLIMMIX can fit the models to data with both fixed and random effects and the response is not necessarily normally distributed (Ying & Liu, 2006).

The estimation of the parameter estimates using this procedure follows likelihood-based methods and the pseudo-likelihood procedure is the default (Schabenberger, 2005). The procedure allows the change of estimation method and specification of covariance structures (Wolfinger & O'connell, 1993).

5.5.1 Application of Generalized Linear Mixed Model to the data

To model the data, we use the PROC GLIMMIX procedure in SAS version 9.4. The response variable is the death of the child under-five and the set of covariates are current age of a child, number of children living, mother's work status, sex of a child, mother's age, weight of a child at birth, region, wealth index, marital status, type of place of residence, educational attainment of mother, birth order number and breast feeding.

In the model statement, the DIST= option states the distribution of the response variable whilst the LINK= option specifies the link function of the PROC GLIMMIX procedure. The binary distribution option with logit link was used in this model. The METHOD= option syntax specifies the parameter and the covariance estimation techniques in the PROC GLIMMIX statement. If the model does not converge, the marginal distribution is approximated by using the Gaussian Quadrature method and the Laplace methods. Both methods produced similar results, with minimal differences in the parameter estimates and standard errors. The random effect were the clusters.

The type 3 tests of fixed effects for the model fitted using Laplace method in GLMs is given in Table 5.1. The F-statistic is used for the significance test for the fixed effects and the p-values that correspond show that most of the effects are important in the model that is fitted when tested at the 5% level of significance.

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Current age of child	4	2718	24.54	<.0001
Mother's age	1	2718	11.02	0.0009
Sex of child	1	2718	0.75	0.3875
Weight of the child at birth	2	2718	3.34	0.0356
Currently breastfeeding	1	2718	0	0.9868
Marital status	1	2718	1.98	0.1591
Wealth index	2	2718	1.47	0.2295
Education attainment	2	2718	3.4	0.0335
Mothers currently working	1	2718	0.43	0.5105
Type of place of residence	1	2718	4.64	0.0314
Region	9	2718	0.8	0.6123
Number of children living	2	2718	23.72	<.0001
Birth order number	2	2718	7.11	0.0008
Child age and currently breastfeeding	4	2718	11.02	<.0001

The effect of sex of child, marital status, mothers work status, weight of child at birth, and region are insignificant as p-values are less than 0.05. The two-way interaction between age of child and currently breastfeeding was significant at five percent level of significance. Table 5.2

shows the type 3 tests of fixed effects for the fitted model using the Laplace method in GLMs for 2014 DHS data.

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
Current age of child	4	2718	24.54	<.0001
Mother's age	1	2718	11.02	0.0009
Sex of child	1	2718	0.75	0.3875
Weight of the child at birth	2	2718	3.34	0.0356
Currently breastfeeding	1	2718	0	0.9868
Marital status	1	2718	1.98	0.1591
Wealth index	2	2718	1.47	0.2295
Education attainment	2	2718	3.4	0.0335
Mothers currently working	1	2718	0.43	0.5105
Type of place of residence	1	2718	4.64	0.0314
Region	9	2718	0.8	0.6123
Number of children living	2	2718	23.72	<.0001
Birth order number	2	2718	7.11	0.0008
Child age and currently breastfeeding	4	2718	11.02	<.0001

The F-statistic is used for the significance test for the fixed effects and the corresponding p-values show that most of the effects are important in the model fitted when tested at the 5% level of significance. The effect of sex of child with a p-value equals 0.3875, currently breastfeeding with a p-value equals 0.9868, marital status with a p-value equals 0.1591, wealth index with a p-value equals 0.2295, mothers work status with a p-value equals 0.5105 and region with a p-value equals 0.6123 are insignificant as the p-values are greater than 0.05. The two-way interaction between age of child and currently breastfeeding (p-value<.0001) was significant at 5%. Table 5.3 show the odds ratios and the parameter estimates obtained using the PROC GLIMMIX procedure. The results obtained were a bit different from those found using PROC LOGISTIC in GLMs. Current age of child and one interaction term are variables that were found to be significant in ordinary logistic are not significant in GLMMs. This could be because of including random effects in the model.

Current age of child was negatively associated with the under 5 child mortality. Mother's age was observed to be significantly associated with child mortality with a p-value <.0001. The odds ratio that corresponds was 1.11 with 95 percent confidence interval (1.074-1.146). The sex of child was not associated with child mortality with a p-value equals 0.0832. The effect of weight of the child that is average and small was observed to be not significantly associated with child mortality (p-value=0.7713 and 0.1698). The covariate, currently breastfeeding, was positively associated with child mortality with a p-value=<.0001. The OR that corresponds was 3.294 with 95% confidence interval (1.21-8.97).

Table 5.1: Estimates and odds ratios (OR) with 95% confidence intervals for the fixed effects. (2009 DHS DATA)

Solutions for fixed effects for 2009 Data							
						95% CI	
Effect	**Estimate**	**SE**	**t-val**	**P>t**	**OR**	**Lower**	**Upper**
Intercept	-6.8415	0.8067	-8.48	<.0001			
Current age of child (ref. 0)							
1 year	-1.7821	0.3632	-4.91	<.0001	0.083	0.054	0.127
2 years	-2.051	0.7312	-2.8	0.0051	0.029	0.013	0.064
3 years	-2.432	0.999	-2.43	0.015	0.019	0.006	0.055
4 years	-4.0007	1.9854	-2.02	0.044	0.004	<0.001	0.036
Mother's age	0.1042	0.01654	6.3	<.0001	1.11	1.074	1.146
Sex of child (ref. Male)							
Female	-0.2644	0.1526	-1.73	0.0832	0.768	0.569	1.035
Weight of the child at birth (ref. Large)							
Average	-0.05981	0.2058	-0.29	0.7713	0.942	0.629	1.41
Small	0.3376	0.2459	1.37	0.1698	1.402	0.865	2.27
Currently breastfeeding (ref. Yes)							
No	3.2616	0.1978	16.49	<.0001	3.294	1.21	8.97
Marital status (ref. Not Married)							
Married	0.04747	0.185	0.26	0.7976	1.049	0.73	1.507
Wealth index (refresh)							
Middle	0.02726	0.2597	0.1	0.9164	1.028	0.618	1.71
Poor	0.6171	0.235	2.63	0.0087	1.854	1.169	2.939
Education attainment (ref.Higher Education)							
No Education	1.4473	0.6982	2.07	0.0382	4.252	1.082	16.712
Up to Secondary Education	1.2327	0.5057	2.44	0.0148	3.431	1.273	9.247
Mothers currently working (ref. Employed)							
Unemployed	-0.2447	0.1649	-1.48	0.1379	0.783	0.567	1.082
Type of place of residence (ref.Urban)							
Rural	0.7553	0.2697	2.8	0.0051	2.128	1.254	3.611

Region (ref. Thaba-Tseka)							
Berea	0.2113	0.3638	0.58	0.5615	1.235	0.605	2.521
Butha-Bothe	0.4182	0.3634	1.15	0.2499	1.519	0.745	3.098
Leribe	0.3951	0.3284	1.2	0.229	1.485	0.78	2.826
Mafeteng	0.8409	0.3373	2.49	0.0127	2.318	1.197	4.491
Maseru	0.2963	0.3444	0.86	0.3897	1.345	0.685	2.642
Mohale's Hoek	0.4075	0.3347	1.22	0.2234	1.503	0.78	2.897
Mokhotlong	0.1416	0.3279	0.43	0.6659	1.152	0.606	2.191
Qacha's-Nek	0.222	0.371	0.6	0.5497	1.249	0.603	2.584
Quthing	0.5155	0.3483	1.48	0.1389	1.674	0.846	3.315
Birth order number (ref. first birth)							
2-4 births	0.7858	0.2683	2.93	0.0034	2.194	1.297	3.713
>4 births	0.7767	0.4929	1.58	0.1151	2.174	0.827	5.715
Number of children living (ref. <2 children)							
2-4 children	-1.9279	0.2692	-7.16	<.0001	0.145	0.086	0.247
>4 children	-2.2151	0.5054	-4.38	<.0001	0.109	0.041	0.294
Significant Interactions Effects							
Current age of child (ref. 0 years) and currently breastfeeding (ref. yes)							
1 years and Not Breastfeeding	-1.4218	0.4332	-3.28	0.001	0.241279	0.10322	0.56399
2 years and Not Breastfeeding	-2.9931	0.8109	-3.69	0.0002	0.050132	0.01023	0.24568

Marital status was not significantly associated with the probability of child mortality with a p-value equals 0.7976. The odds ratio that corresponds was 1.049 with 95% confidence interval (0.73-1.507). The effect of wealth index middle was not significantly associated with the probability of child mortality with a p-value equals 0.9164. The odds ratio that corresponds was 1.028 with 95% confidence interval (0.618-1.71). The effect of wealth index poor was observed to be positively associated with the probability of child mortality with a p-value equals 0.0087. The odds ratio that corresponds was 1.854 with 95% confidence interval (1.169- 2.939). The odds of dying for a child who's from a poor family were 1.854 times the odds of dying for a child who's from a rich family. The effect of no education was observed to be positively associated with child mortality with a p-value equals 0.0382. The odds ratio that corresponds was 4.252 with 95% confidence interval (1.082-16.712). The odds of dying of a child from a mother who has no educations is 4.252 times the odds of dying of a child from a mother with higher education. The effect of up to secondary education was positively associated with the probability of child mortality with a p-value equals 0.0148. The odds ratio that corresponds was 3.431 with 95% confidence interval (1.273-9.247). The odds of dying for a child who is from a mother with up to secondary education were 3.431 times the odds of dying for a child who's from a mother with higher education. The effect of type of residence that is rural was observed to be positively associated with child mortality with a p-value equals 0.0051. The odds ratio that corresponds was 2.128 with 95% confidence interval (1.254 -3.611). The odds of dying for a child who is from a rural area were 2.128 times the odds of dying for a child who was from the urban area.

The mother's employment status was found not significantly associated with the probability of child mortality. The effect of Region Mafeteng was the only region significantly associated with the probability of child mortality with a p-value equals 0. 0127. The odds ratio that corresponds was 2.318 with 95% confidence interval (1.197 -4.491). The odds of dying for a child who was from Mafeteng region were 2.318 times the odds of dying for a child who was from Thaba-Tseka region. The effect of childbirth order number that is between two and four was significantly associated with child mortality with a p-value equals 0.0034. The odds ratio that corresponds was 2.194 with 95% confidence interval (1.297-3.713). The number of children living within two to four and the number of children living which is more than four was negatively associated with child mortality (p-value=<.0001). The odds ratios that correspond were 0.145 with 95% confidence interval (0.086 ; 0.247) and 0.109 with 95% confidence interval (0.041-0.294) respectively. The odds of dying for a child from a mother with two to four children living were 0.145 times the odds of dying for a child from a mother with less than two children alive. The odds of dying for a child from a mother with more than four children alive were 0.109 times the odds of dying for a child from a mother with less than 2 children alive.

The two-way interaction effects for not breastfeeding and child who is 1 year is negatively associated with the child mortality with a p-value equals 0.001. The odds ratio that corresponds was 0.2413 and 95% confidence interval (0.1032-0.564). The odds of dying for a child who was 1 year and from a mother who does not breastfeed was 0.2413 times the odds of dying for a child who was 0 year and from a mother who breastfeeds. The two-way interaction effects for breastfeeding ("No") by current age of child 2 year is negatively associated with the child mortality with a p-value equals 0.0002. The odds ratio that corresponds was 0.050 with 95% confidence interval(0.0102-0.246). The odds of dying for a child from a mother who does not breastfeed by age of child 2 year was 0.05 times the odds of dying for a child from a mother who breastfeeds by age of child 0 year. The two-way interaction effects for not breastfeeding a child at age 1 and 2 was not significant and the analysis only shows significant effects for two-way interaction. Table 5.4 show the same table results as 5.3 but for the 2014 DHS DATA. The results obtained were

Table 5.2: Estimates and odds ratios (OR) with 95% confidence intervals for the fixed effects. (2014 DHS DATA)

Solutions for fixed effects for 2014 Data							
						95% CI	
Effect	**Estimate**	**SE**	**t-val**	**P>t**	**OR**	**Lower**	**Upper**
Intercept	-8.0788	1.0382	-7.78	<.0001			
Current age of child (ref. 0)							
1 year	-1.816	0.6271	-2.9	0.0038	0.071	0.036	0.14
2 years	-0.2463	0.6735	-0.37	0.7146	0.061	0.027	0.136
3 years	-4.1867	4.641	-0.9	0.3671	0.004	<0.001	0.399
4 years	-10.6499	110.21	-0.1	0.923	<0.001	<0.001	.
Mother's age	0.07794	0	∞	<.0001	1.081	.	
Sex of child (ref. Male)							
Female	0.1872	0.2083	0.9	0.3687	1.206	0.802	1.814
Weight of the child at birth (ref. Large)							
Average	0.227	0.2848	0.8	0.4254	1.255	0.718	2.193
Small	0.803	0.3361	2.39	0.0169	2.232	1.155	4.315
Currently breastfeeding (ref. Yes)							
No	4.0073	0.2813	14.25	<.0001	1.03	<0.001	>999.999
Marital status (ref. Not Married)							
Married	0.3855	0.2591	1.49	0.1369	1.47	0.885	2.444
Wealth index (refresh)							
Middle	0.03088	0.308	0.1	0.9201	1.031	0.564	1.887
Poor	0.4545	0.305	1.49	0.1363	1.575	0.866	2.865
Education attainment (ref.Higher Education)							
No Education	1.5738	1.2359	1.27	0.203	4.825	0.428	54.44
Up to Secondary Education	1.4226	0.5325	2.67	0.0076	4.148	1.46	11.784
Mothers currently working (ref. Employed)							
Unemployed	0.1766	0.2544	0.69	0.4877	1.193	0.725	1.965
Type of place of residence (ref.Urban)							
Rural	0.6333	0.2956	2.14	0.0322	1.884	1.055	3.364
Region (ref. Thaba-Tseka)							
Berea	0.2141	0.5051	0.42	0.6716	1.239	0.46	3.335
Butha-Bothe	-0.0598	0.5085	-0.12	0.9064	0.942	0.348	2.553
Leribe	0.8054	0.4842	1.66	0.0964	2.238	0.866	5.783
Mafeteng	0.7434	0.513	1.45	0.1474	2.103	0.769	5.75
Maseru	0.2113	0.4887	0.43	0.6655	1.235	0.474	3.22
Mohale's Hoek	0.3694	0.4824	0.77	0.4438	1.447	0.562	3.726
Mokhotlong	0.5744	0.4867	1.18	0.2381	1.776	0.684	4.612

Qacha's-Nek	0.4051	0.5611	0.72	0.4703	1.5	0.499	4.506
Quthing	0.6311	0.5024	1.26	0.2091	1.88	0.702	5.034
Birth order number (ref. first birth)							
2-4 births	1.2401	0.3583	3.46	0.0005	3.456	1.712	6.978
>4 births	1.9601	0.6386	3.07	0.0022	7.1	2.03	24.837
Number of children living (ref. <2 children)							
2-4 children	-2.4474	0.3647	-6.71	<.0001	0.087	0.042	0.177
>4 children	-3.2703	0.7004	-4.67	<.0001	0.038	0.01	0.15
Significant Interactions Effects							
Current age of child (ref. 0 years) and currently breastfeeding (ref. yes)							
1 years and Not Breastfeeding	-1.6644	0.6933	-2.4	0.0164	0.189304	0.048642	0.736731
2 years and Not Breastfeeding	-5.1013	0.8093	-6.3	<.0001	0.006089	0.001246	0.029746

the same as those obtained using PROC LOGISTIC in GLMs. The current age of child that is 1 year was established as negatively associated with the child mortality. Mother's age was significantly associated with child mortality with a p-value=<.0001.The odds ratio that corresponds was 1.081. Sex of child was insignificantly associated with child mortality with a p-value equals 0.3687. The odds ratio that corresponds was 1.206 with 95% confidence interval: 0.802-1.814. The effect of weight of the child that is average was not significantly associated with child mortality with a p-value equals 0.4254. The effect of weight of the child that is small was observed as positively associated with child mortality with a p-value equals 0.0169. The odds ratio that corresponds was 2.232 with 95% confidence interval (1.155 -4.315). Currently breastfeeding was positively associated with child mortality with a p-value=<.0001. The odds ratio that corresponds was 1.03.

Marital status was not significantly associated with child mortality with a p-value equals 0.1369. The odds ratio that corresponds was 1.47 with 95% confidence interval (0.885-2.444). The effect of wealth index middle was found not significantly associated with child mortality with a p-value equals 0.9201. The odds ratio that corresponds was 1.031 with 95% confidence interval (0.564-1.887). The effect of wealth index poor was found not significantly associated with child mortality with a p-value equals 0.1363. The odds ratio that corresponds was 1.575 with 95% confidence interval (0.866 -2.865). The effect of no education was found not significantly associated with child mortality with a p-value equals 0.203. The odds ratio that corresponds was 4.825 with 95% confidence interval (0.428 -54.44). The effect of up to secondary education was observed as positively associated with child mortality with a p-value equals 0.0076. The odds ratio that corresponds was 4.148 with 95% confidence interval (1.46-11.784). The effect of type of residence that is rural was positively associated with child mortality with a p-value equals 0. 0322. The odds ratio that corresponds was 1.884 with 95% confidence interval (1.055 -3.364).

The mother's employment status was not significantly associated with child mortality. The effect of Region was not significantly associated with child mortality. The effect of childbirth order number that is between two and four was found to be significantly associated with child mortality with a p-value=<0.0001. The odds ratio that corresponds was 2.194 with 95% confidence interval (1.297-3.713). The effect of childbirth order number that is between two and four was observed as positively associated with child mortality with a p-value equals 0.0005. The odds ratio that corresponds was 3.456 with 95% confidence interval (1.712-6.978). The effect of childbirth order number that is more than four was observed as positively associated with child mortality with a p-value equal s 0.0022. The odds ratio that corresponds was 7.1 with 95% confidence interval (2.03 -24.837).

The number of children living within two to four and the number of children living which is more than four was observed as negatively associated with child mortality with a p-value=<.0001. The odds ratios that correspond were 0.087 with 95% CI (0.042 – 0.177) and 0.038 with 95% confidence interval (0.01-0.15) respectively. The odds of dying for a child from a mother with two to four children living were 0.087 times the odds of dying for a child from a mother with less than two children alive. The odds of dying for a child from a mother with more than four children alive were 0.038 times the odds of dying for a child from a mother with less than 2 children alive.

The two-way interaction effects for breastfeeding ("No") by current age of child 1 year is negatively associated with the child mortality with a p-value equals 0.0164. The odds ratio that corresponds was 0.1893 with 95% confidence interval (0.0486-0.7367). The odds of dying for a child from a mother who does not breastfeed by age of child 1 year was 0.1893 times the odds of dying for a child from a mother who breastfeeds by age of child 0 year. The two-way interaction effects for breastfeeding ("No") by current age of child 2 year is negatively associated with the child mortality with a p-value= <. 0001. The odds ratio that corresponds was 0.0061 with 95% confidence interval (0.00125 0.0297). The odds of dying for a child from a mother who does not breastfeed by age of child 2 year was 0.0061 times the odds of dying for a child from a mother who breastfeeds by age of child 0 year. Figure 5. and Figure 5.2 shows adjusted comparison of breastfeeding by current age of child interaction least-square means for multiplicity based on 2009 and 2014 DHS DATA. The lines that represent the significant difference between the least-square means of the level of breastfeeding by mother's age interaction effects are the ones centered. The lines that cross the 45-degree line show that the child mortality is not significant between corresponding categories.

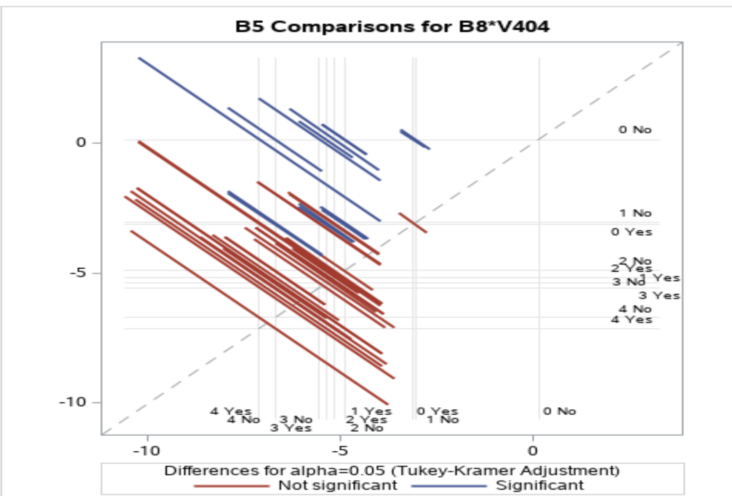

Figure 5.1: Diffogram for breastfeeding by current age of child.(2009 DHS DATA).

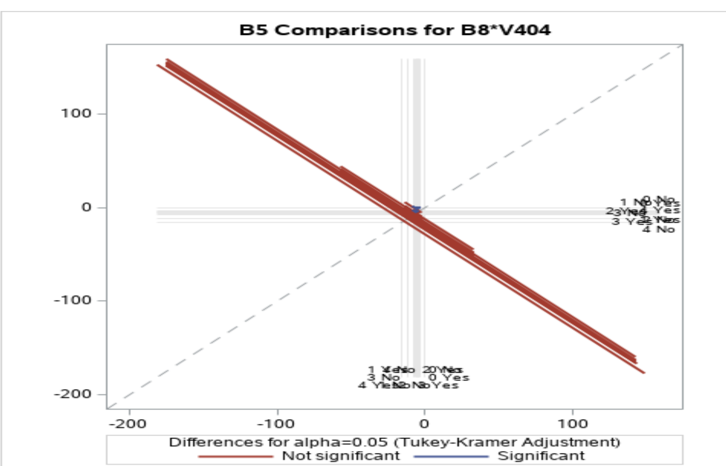

Figure 5.2: Diffogram for breastfeeding by current age of child.(2014 DHS DATA).

The average of breastfeeding by current age of child interaction effect (on logit scale) is -4.4103 and -6.5711 as given by Figure 5.3 and Figure 5.4.

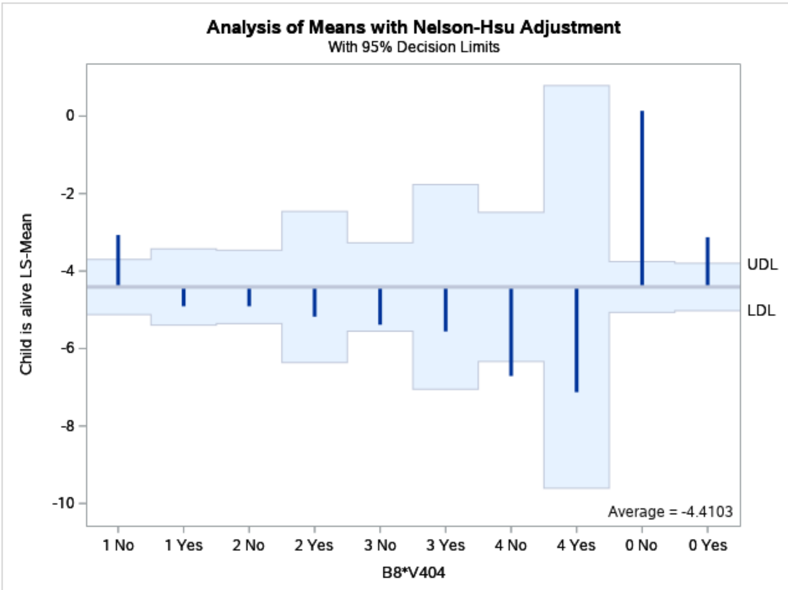

Figure 5.3: Analysis of means for breastfeeding by current age of child interaction effects(2009).

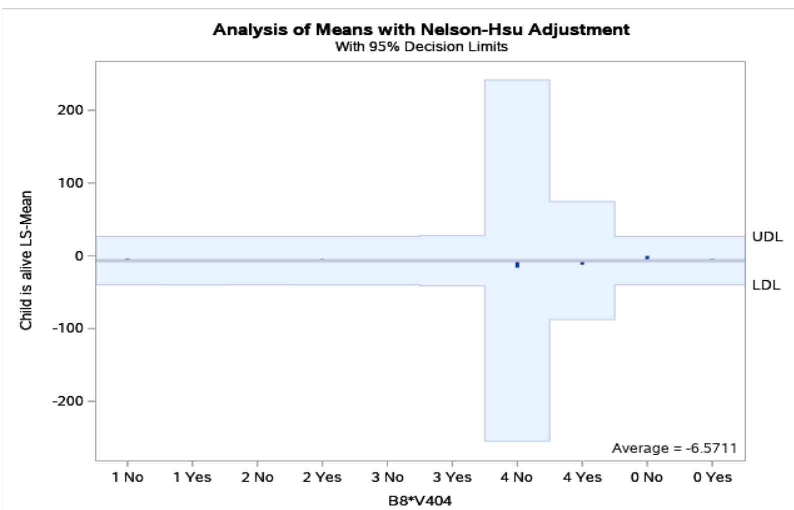

Figure 5.4: Analysis of means for breastfeeding by current age of child interaction effects(2014).

From this figure the differences of means of levels with the vertical lines that crosses 95% decision limits mean that they are significant. The adjusted comparison of region least-squares means for multiplicity are shown in Figure 5.5 and Figure 5.6. All the regions are not significant. The average region effect (on logit scale) is -4.5922 and -6.6807, as given in Figure 5.7 and Figure 5.8.

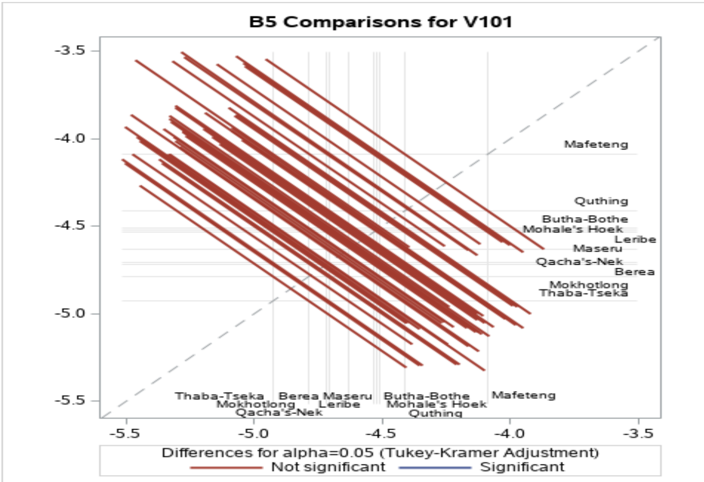

Figure 5.5: Diffogram for Region effect(2009).

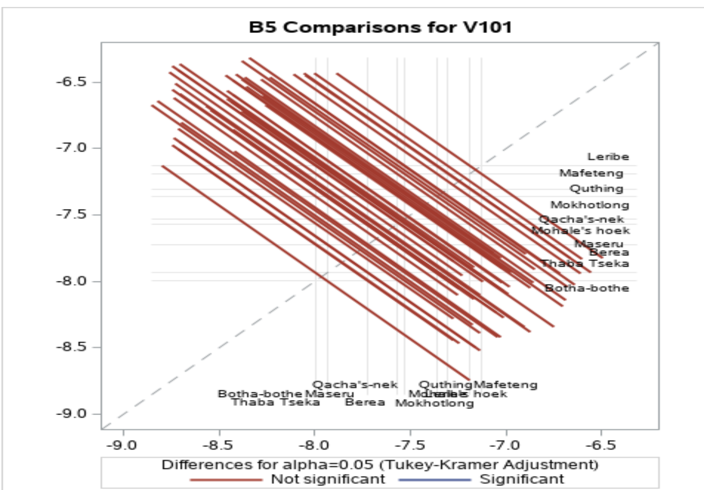

Figure 5.6: Diffogram for Region effect(2014).

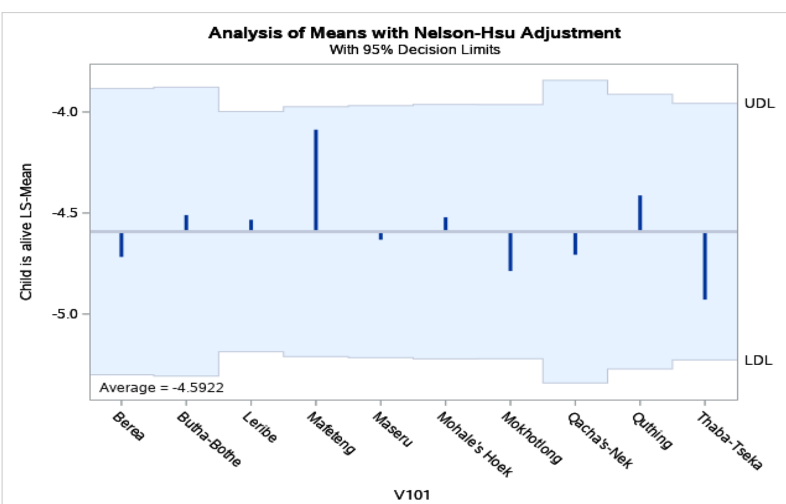

Figure 5.7: Analysis of means for Region effect(2009).

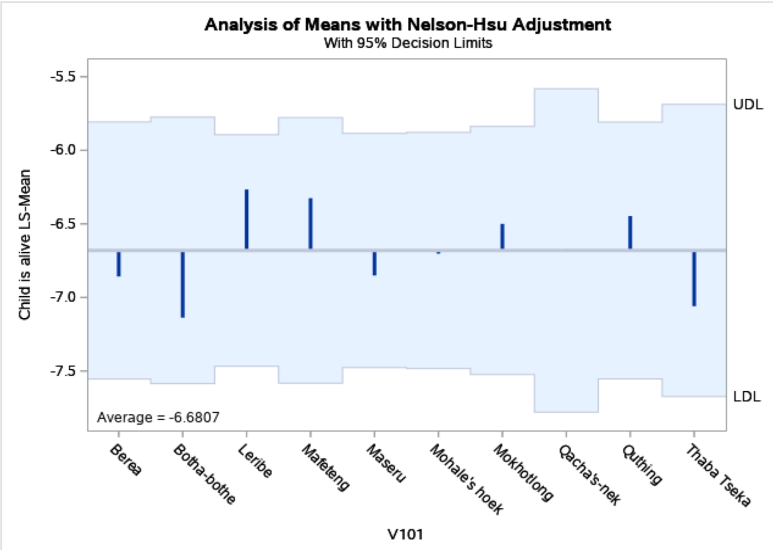

Figure 5.8: Analysis of means for Region effect(2014).

5.5.2 Summary of Generalized Linear Mixed Models

The study shows that most of these factors are important in explaining the probability of under-five child mortality. The factors that had an effect are current age of a child, number of children living, mother's age, educational attainment of mother, wealth index, type of place of residence, birth order number and breastfeeding. Furthermore, the two-way interaction that is associated with the probability of under-five mortality is breastfeeding by current age of child. The next chapter introduces the generalized additive models.

Chapter 6

6 Semi Parametric Regression Approach

6.1 Introduction

In the previous chapters, the relationship between child mortality and the predictor variables have been studied using parametric methods. In these approaches, the functional form of the model is assumed known prior to modelling and the interpretation of the parameters is easy. However, the parametric methods suffer from bias that is introduced during estimation. Therefore there is a need for nonparametric modelling that assumes an unknown functional form of the model, before modelling (Härdle et al., 2012). Nonparametric regression is a form of regression analysis in which none of the predictors take predetermined functional forms with the response but are constructed according to the information derived from the data (Eubank, 1999).

The parametric models do not give a reasonable picture of what is occurring pragmatically although they are easy to understand and work with. On the other hand, non-parametric methods might represent the data very well but not suited analytically. Therefore, it is advisable to use semiparametric models that combine the nonparametric and parametric model. Semiparametric models permits you to have a model that is understandable and offers a fair representation of the complexity that is involved (Burman & Chaudhuri, 2012; Härdle et al., 2012). The combination of parametric and non-parametric methods is much more powerful than using only one of the two methods (Bush & MacEachern, 1996).

The regression model that describes the relationship between the response variable and the covariates may be written as:

$$Y_i = f_1(x_1, \beta_1) + f_1(x_2, \beta_2) + \epsilon \qquad (6.1)$$

parameters that need to be estimated are given by β_1 and β_2 while the random error term is ϵ, and $f_1(.)$ and $f_2(.)$ are functions that explain the relationship.
A semiparametric model can take many structures. One is a form of regression analysis where some of the forms taken by predictors is not predetermined and another takes known forms with

the response (Mahmoud, 2019). In the model above, f_1 may be known and f_2 is unknown. The model can be written as:

$$Y = \beta_0 + \beta_1 x_1 + f(x_2) + \epsilon \qquad (6.2)$$

Single index model (SIM) is another way of modeling semiparametric regression and has a lot of applications and is widely studied (Yu & Ruppert, 2002). It takes the form:

$$\mathrm{Y}=f(X\beta) + \epsilon \qquad (6.3)$$

f is an unknown function with X = $(x_1, \ldots, x_k)$ being a n × k matrix of regressors values, ϵ the error that satisfies $E(\epsilon\ |\mathrm{X})$ and β is a k × 1 vector of parameters.

In comparison to parametric models the Single index model is more flexible. SIM assumes the link between the explanatory variables and the mean response is unknown and then estimates it non-parametrically (Ahn & Powell, 1993).

This chapter makes practical use of this form of regression in the generalized additive effects model (GAM) in identifying the factors associated with the probability of child mortality in Lesotho.

6.2 Generalized Additive Models

The generalized additive model (GAM) is an example of the semiparametric regression models. It is a generalization of the GLM for modelling non-gaussian data and an extension of the nonparametric additive model (Vickers, 2005). The GAM can be applied when dealing with standard continuous response regression, count data, dichotomous response, survival data and time series data (Guisan et al., 2002). GAMs are appropriate for exploring the data set and visualizing the relationship between the dependent and independent variables (Liu, 2008).

The logistic regression models the effects of covariates x_j in terms of the linear predictor of the form $\sum x_j\beta_j$ where the β_j are the model parameters. The generalized additive models generalize the GLMs by replacing $\sum x_j\beta_j$ with $\sum s_j(x_j)$ where s_j is unspecified (nonparametric) function. GAMs originally developed by (Hastie & Tibshirani, 1990) are semi-parametric extensions of

GLMs; they rely on the assumption that functions are additive and that added components are smooth. This function is estimated in a flexible manner using a cubic smoother (Hastie & Tibshirani, 1990). The robustness of GAMs is the ability to handle non-linear and non-monotonic association between dependent and independent variables (Guisan et al., 2002).

6.2.1 Additive Models

The linear regression model assumes the expected value of the response variable Y has a linear form the following type.

$$E(Y) = f(X_1, \ldots\ldots, X_p) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p \tag{6.4}$$

Considering a sample of values for X and Y, estimates of $\beta_0, \beta_1, \ldots, \beta_p$ are found by the least square method. The additive model generalizes the linear model by modeling the expected value of Y as:

$$E(Y) = f(X_1, \ldots\ldots, X_p) = s_0 + s_1(X_1) + \ldots + s_p(X_p) \tag{6.5}$$

where $s_i(X)$, $i = 1,2, \ldots, p$ are smooth functions. The most used method for estimating the GAMs is the backfitting algorithm. It enables one to fit the additive model using any of the regression fitting mechanisms (Binder & Tutz, 2008). To obtain the best fit model for the data, one needs to find an approximate estimate of the smooth function (Silverman, 1985) .This is usually the main concern in additive regression and hence the estimate can be obtained through smoothing. We first look at the methodology before we consider the form of the logistic regression in the GAMs models setting.

6.3 Smoothing

6.3.1 What is a Smoother?

A tool that gives a summary of the trend of a response measurement Y as a function of predictor measurements $X_1, X_2; \ldots; X_p$ is called a smoother. It gives an estimate of the trend that is less variable than Y. The most crucial attribute of smoother is its non-parametric nature. The estimate produced by a smother is a smooth (Liu, 2008; Wood et al., 2016) and doesn't assume an inflexible form for the dependence of Y on $X_1, X_2; \dots; X_p$.

6.3.2 Cubic Smoothing Splines

First consider the simplest smooth function, where the model comprises of one smooth function of one covariate:

$$Y_i = S(x_i) + \epsilon_i \tag{6.6}$$

where Y_i is the response variable, x_i is the covariate, $S(.)$ is the smooth function and ϵ_i are the error terms such that $\epsilon_i \sim N(0, \sigma^2)$ (Rice & Rosenblatt, 1983).

$S(x_i)$ defines the regression function of y on x, a nonparametric smooth function that must be estimated. To minimize the penalized sum of squares (PSS) is the optimization problem and is stated in this below:

$$\sum \left(y_i - S(x_i)\right)^2 + \lambda \int_a^b [S''(x)]^2 \, dx \tag{6.7}$$

the fixed constant is λ, and $a \leq x_1 \leq \ldots \leq x_n \leq b$. It is assumed (a, b) includes all possible range (Rice & Rosenblatt, 1983; Silverman, 1984).

The smoothing parameter is defined by λ and it controls the trade-off between the curve smoothness and proximity to the values of y. This is usually estimated by restricted marginal likelihood which exploits the link between Bayesian estimation and spline smoothing or by GCV. Notice that when $\lambda \rightarrow 0$ (no smoothing), the solution is an interpolating function and as $\lambda \rightarrow \infty$ (infinite smoothing), the estimate converges to a linear least square estimate (Friedman & Stuetzle, 1982; Pollock, 1993).

Cubic smoothing splines in a simple setting is described below. Suppose that we have a scatterplot of points (x_i, y_i) shown in Figure 6.1.

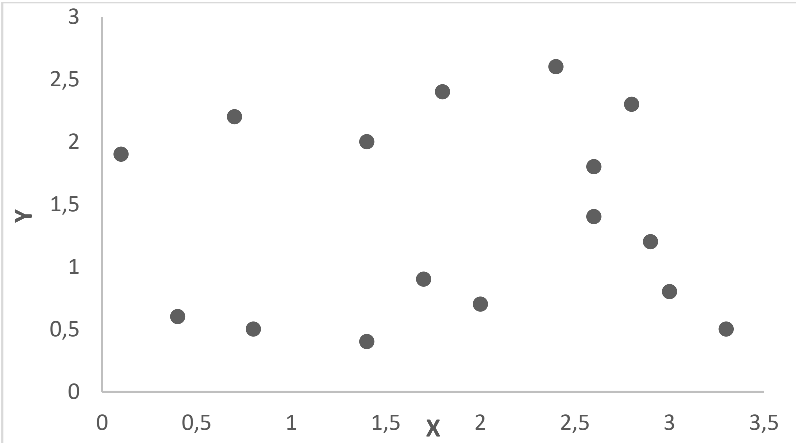

Figure 6.1

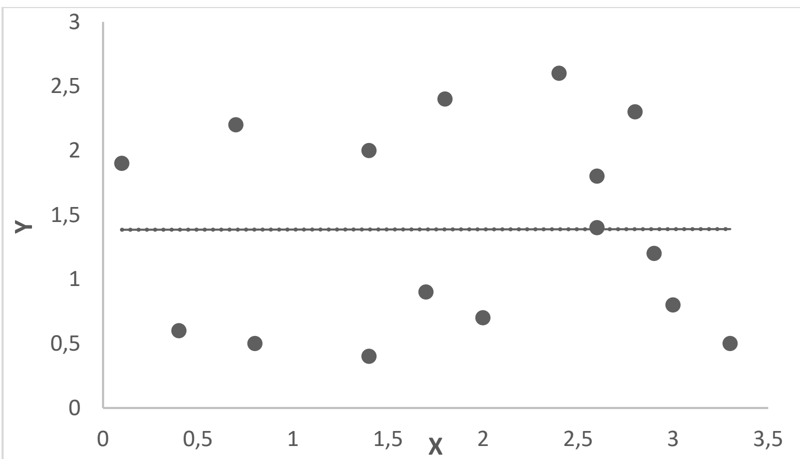

Figure 6.2

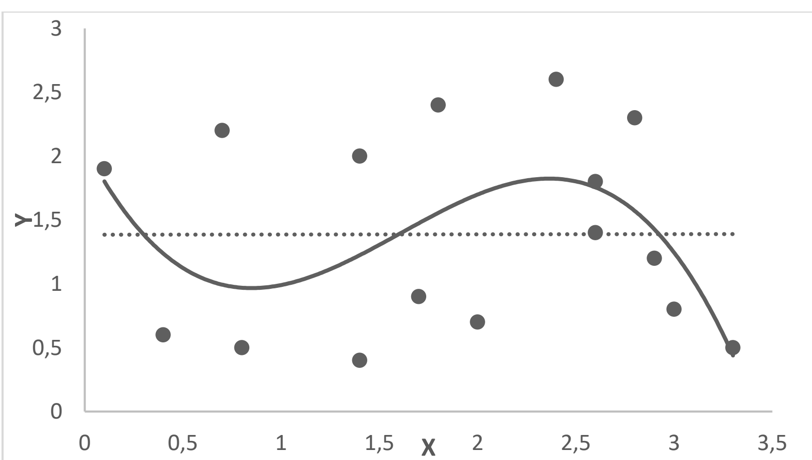

Figure 6.3

Figure 6.1 displays a scatterplot of an outcome that measures *y* by plotting it against an independent variables *x*. In Figure 6.2, the least square method was used to fit the straight line. In Figure 6.3, to describe the trend of *y* on *x* ,a cubic spline is added on. The cubic smoothing spline describes the trend of *y* as a function of *x* better compared to the LS method (Liu, 2008).

6.4 Choosing Smoothing Parameters λ

It is crucial to select a smoothing parameter wisely in order to minimize cubic spline smoother under consideration. The key in this selection process is choosing how much smoothing to do. This is no different than when you specify different parametric forms for an explanatory variable, in that one is trying to specify the correct functional form. If the smoothing parameter is much bigger, then the data will be over smoothed, but if it is too low then the data will be under smoothed (Hurvich et al., 1998). There are other useful approaches when the scale parameter is known than trying to minimize expected mean square error which results into estimation by Un-Biased Risk Estimation. Attempting to minimize the prediction error when the scale parameter is unknown leads to ordinary cross validation or generalized cross validation (Hall & Titterington, 1987).

6.4.1 Average Mean Square and Predictive Square Error

We must select the smoothing parameter, λ, in order to minimize the cubic spline smoother. For this, we don't need to minimize the Mean Square Error at each covariate x_i, but instead we must put attention on a global measure such as Average Mean Square Error (AMSE) (Wood, 2000). The average mean square error is given by:

$$AMSE(\lambda) = \frac{1}{n}\sum_{i=1}^{n} \left[\hat{S}_\lambda(x_i) - S(x_i)\right]^2 \qquad (6.8)$$

where Y_i=$S(x_i)$+ϵ_i and $\hat{S}_\lambda(x_i)$ is an estimator of $S(x)$. The Average Predictive Square Error (PSE) differs from AMSE by only a constant function, σ. The APSE is given by:

$$APSE(\lambda) = \frac{1}{n} \sum_{i=1}^{n}\left[Y^*{}_i - \hat{S}_\lambda(x_i)\right]^2 \qquad (6.9)$$

where $Y^*{}_i$ is a new observation at x_i, that is $Y^*{}_i = S(x_i) + \epsilon^*{}_i$, $\epsilon^*{}_i$ is independent of $\epsilon_i's$, and $E(\epsilon^*{}_i)=0$ (Hoover et al., 1998; Liu, 2008).

6.4.2 Cross Validation(CV)

Cross-Validation is a statistical approach of partitioning a sample of data into two subsets. This method is inefficient unless the sample is large (Picard & Cook, 1984). The whole idea behind cross-validation is recycling data by exchanging the roles of training and test samples in cross validation. We are more interested in the (LOOCV)Leave-one-out Cross-Validation for simplification of the data set (Wang, 2012). This works by dropping the point (x_i, y_i) out one by one as the testing set and approximating the smooth at x_i depending on the remaining n-1 points (Hastie & Tibshirani, 1987). The CV can be used in selecting λ by minimizing:

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left\{y_i - \hat{S}_\lambda{}^{-i}(x_i)\right\}^2 \qquad (6.10)$$

where, $\hat{S}_\lambda{}^{-i}(x_i)$ illustrates the fit at x_i which is calculated by leaving out the i^{th} data point. Minimizing $CV(\lambda)$ is equivalent to minimizing $APSE(\lambda)$, we can use $CV(\lambda)$ for smoothing selection (Wood, 2006; Xiang & Wahba, 1996).

6.4.3 Generalized Cross Validation(GCV)

The Generalized Cross Validation is another method of selecting λ (Golub et al., 1979). This approach has been used widely in a lot of nonparametric regression approaches as a point of reference to select the smoothing parameters (Marcotte, 1995). The Generalized Cross Validation function approximates the expected prediction error. The model with the best prediction ability is deduced by the model selected by the GCV function (Efron, 2004).

This method is complex to compute but previous literatures have discussed in detail the algebra for this approach (Alvarez Meza et al., 2012; Liu, 2008). The GCV is given by:

$$GCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i - \hat{S}_\lambda(x_i)}{1 - \frac{trS}{n}}\right]^2 \qquad (6.11)$$

GCV is a weighted version of CV (Golub et al., 1979).

6.4.4 Degrees of Freedom of a Smoother

Using degrees of freedom is another way of expressing the smoothness that's required of the function other than in terms of smoothing parameter. The degrees of freedom of a smoother *df(smoother)*, other times referred to as the effective number of parameters, indicates the amount of smoothing (Buja et al., 1989; Wood et al., 2016). With a linear smoother given by S_λ, the degrees of freedom *df* is defined as *df(smoother)=tr(* S_λ*)*. The smoothing parameter is the crucial determining factor of the degrees of freedom. More smoothing means higher span or fewer *df(smoother)*. The *df(smoother)* doesn't have to be an integer.

In SAS procedure, one can select the value of a smoothing parameter through stating the degrees of freedom of a smoother. This indicates the amount of smoothing. Fewer degrees of freedom of the smoother result when there is more the smoothing (Liu, 2008).

The proc GAM applies thin-plate smoothing spline for bivariate smoothing components and applies the local regression methods and B-spline for univariate smoothing components. The GAM procedure automatically chooses the smoothing parameter by GCV as the model chosen by the GCV function is considered to have the best ability to do prediction.

6.5 Fitting the Generalized Additive Models

The general method of fitting additive models will be outlined in this section. The two major pieces are the back-fitting and local scoring algorithms. The GAMs framework is based on back-fitting with linear smoothers, there are limitations because of the difficulty that results in back-fitting in the selection of a model and inference.

"The back-fitting algorithm cycles through the individual terms in the additive model and updates each using an appropriate smoother. This is done by smoothing partial residual that are suitably defined" (Buja et al., 1989). The local scoring algorithm is similar to the Iterative Reweighted Least Squares used to fit GLM or the Fisher-scoring algorithm. Each iteration gives a new working response and weights, which are handed to a weighted back-fitting algorithm, which yields a new additive predictor. There are many other ways to approach estimating the additive models. "The back-fitting algorithm is one that can generally fit an additive model with any regression-type fitting mechanisms" (Buja et al., 1989; Jain et al., 2017).

6.5.1 Back-fitting Algorithm

Define the partial residuals as

$$R_k = Y - S_0 - \sum_{j \neq k} S_j(x_j) \tag{6.12}$$

Then $E(R_k/x_k) = S_k(x_k)$. This observation gives us a way for estimating each smoothing function $S_k(.)$. With an observation (x_i, y_i) , a criterion like the penalized sum of squares can be stated for this problem, which is

$$\sum_{k=1}^{p} \left(y_k - S_0 - S_k(x_k)\right)^2 + \sum_{k=1}^{p} \lambda_k \int [S_k''(t_k)]^2 \, dt_k \tag{6.13}$$

The iterative procedure that results is called the back-fitting algorithm (Friedman & Stuetzle, 1981). The formulae that follows is from (Hastie & Tibshirani, 1987; Hastie & Tibshirani, 1990).

The unweighted form of the back-fitting algorithm is:

1. Initialize:

$$S_0 = E(Y).$$

$${S_1}^1 = {S_2}^1 = \ldots . = {S_p}^1 = 0; \text{ m} = 0$$

2. Iterate:

$$\text{m} = m + 1 \text{ ;}$$

$$\text{for } j = 1 \text{ to } p.$$

Find:

$$R_j = Y - S_0 - \sum_{k=1}^{j-1} S_k{}^m(x_k) - \sum_{k=j+1}^{p} S_k{}^{m-1}(x_k)\ ;$$

$$S_j{}^m = E(R_j|x_j)$$

3. Calculate:

$$\text{RSS} = \frac{1}{n}\left\| y - S_0 - \sum_{j=1}^{p} S_j{}^m(x_j)\right\|^2$$

Until the convergence criterion is reached or fails to decrease. $S_j{}^m(.)$, denotes the estimate $S_j(.)$ at the *m*th iteration. With many smoothers, RSS does not get any larger at any step. This means that the algorithm will always converge.(Hastie & Tibshirani, 1987; Hilbe, 1993; Xiang, 2001).

6.5.2 The General Local Scoring Algorithm

1. Initialise:

   $S_i = g(E(y))\ ;\ S_1{}^0 = S_2{}^0 = \ldots = S_p{}^0 = 0, m = 0$

2. Iterate:

   $\mathrm{m} = m + 1;$

   Form the weights w, mean μ, predictor η, and the adjusted dependent variable:

$$Z_i = \eta_i + (y_i - \mu_i)\left(\frac{d\eta_i}{d\mu_i}\right)$$

   here, $\eta^{m-1} = S_0 + \sum_{j=1}^{p} S_j{}^{m-1}(x_{ij})$

$$\text{therefore } \eta^{m-1} = g(\mu^{m-1})$$

thus

$$\mu^{m-1} = g^{-1}(\eta_i)$$

construct weights

$$w_i = \left(\frac{d\eta_i{}^{m-1}}{d\mu_i{}^{m-1}}\right)^2 V_i{}^{-1}$$

where $V_i = var(Y_i)$. Using the back-fitting algorithm with weights w**,** fit a weighted additive model to Z_i to get an estimated function $S_j{}^m$, additive predictor η^m, and fitted value $\mu_i{}^m = p_i$.

3. Until: Repeat until the deviance fails to decline or the convergence criterion is satisfied.

The following condition is used in the GAM procedure as the convergence criterion for local scoring:

$$\frac{\sum_{i=1}^{n} w_i \sum_{j=1}^{p} \left(S_j{}^{m-1}(x_{ij}) - S_j{}^m(x_{ij})\right)^2}{\sum_{i=1}^{n} w_i \sum_{j=1}^{p} \left(1 + S_j{}^{m-1}(x_{ij})\right)^2} \leq \epsilon^s$$

where $\epsilon^s = 10^{-8}$ by default (Hastie & Tibshirani, 1987).

Provided the initial estimations of $\eta(x)$, a first order Taylor series expansion with the fisher scoring method, will result in the estimate of:

$$\eta^{est}(x) = \eta^{given}(x) + \delta \qquad (6.14)$$

$$\delta = \frac{Score\ function}{Expected\ information\ matrix}$$

$$= \frac{\frac{dL}{d\eta}}{E\left(-\frac{d^2L}{d\eta^2}/x\right)}$$

$$= E\left(\eta(x) - \frac{\frac{dL}{d\eta}}{E\left(-\frac{d^2L}{d\eta^2}/x\right)}/x\right) \qquad (6.15)$$

Using the chain rule:

$$\frac{dL}{d\eta} = \frac{dL}{d\mu} \cdot \frac{d\mu}{d\eta}$$

$$L = ln \prod_{i=1}^{p} \mu_i{}^{y_i}(1-\mu_i)^{1-y_i}$$

$$\frac{dL}{d\mu_i} = \mu_i \cdot \frac{1}{\mu_i} + (1 - y_i)\frac{-1}{1 - \mu_i}$$

$$= \frac{(1-\mu_i)\mu_i - \mu_i(1-y_i)}{\mu_i(1-\mu_i)} = \frac{y_i - \mu_i}{(1-\mu_i)\mu_i} \tag{6.16}$$

Since $var(Y_i) = E(Y_i{}^2) - \left(E(Y_i)\right)^2 = 1^2\mu_i + 0^2(1-\mu_i)(-\mu_i) = \mu_i(1-\mu_i)$

So,

$$\frac{dL}{d\eta} = \frac{dL}{d\mu} \cdot \frac{d\mu}{d\eta}$$

$$\frac{dL}{d\eta} = (y - \mu) \cdot V^{-1} \cdot \frac{d\mu}{d\eta}$$

$$\frac{d^2L}{d\eta^2} = (y - \mu) \cdot \frac{d}{d\eta} \cdot \left(V^{-1}\frac{d\mu}{d\eta}\right) - \left(\frac{d\mu}{d\eta}\right)^2 V^{-1}$$

And therefore,

$$E\left(\frac{d^2L}{d\eta^2}/x\right) = -\left(\frac{d\mu}{d\eta}\right)^2 V^{-1}$$

$$\eta^{est}(x) = E\left(\eta(x) + (y - \mu) \cdot \frac{d\eta}{d\mu}/x\right) \tag{6.17}$$

Then replace the conditional estimations by smoothers

$$\eta^{est}(x) = smoother\left(\eta(x) + (y - \mu) \cdot \frac{d\eta}{d\mu}/x\right) \tag{6.18}$$

6.6 Spline Bases and Penalties

The penalized spline smoothing goes back to O'Sullivan in 1986, but it was (Eilers & Marx, 1996) who first introduced the combination of difference penalties called P-splines and B-splines (Wand & Ormerod, 2008). The term B-spline is abbreviated for basis spline and was created by Isaac Jacob Schoenberg (Al-Rawi et al., 2010). A B-spline of order n is a piecewise polynomial function of degree n-1 in a variable x. Knots is the place where the pieces meet. Spline functions and their derivatives could be continuous, depending on the multiplicities of the knots (De Boor et al., 1978). B-splines can have the same subset of their knots, but two B-splines are equal when specified over the same knots as a B-spline is uniquely defined by its knots.

"The B-spline of degree q contains q + 1 polynomial pieces of each of degree q; the polynomial pieces come together at q inner knots. At the joining points , the derivatives up to order q − 1 are continuous. The B-spline is positive on a domain spanned by q + 2 knots; everywhere else is zero except at the boundaries where it overlaps with 2q polynomial pieces of its neighbors. For given x, q + 1 B-splines are non-zero" (Marx & Eilers, 1998). The concluding form of the curve generated by a B-spline of degree q is given by

$$S(x) = \sum_{i=1}^{n} a_i B_{i,q}(x) \tag{6.19}$$

where, n, is the total number of B-spline basis being used. When the data is being interpolated, estimation of the curve is done with the least squares to obtain the optimum values of the control points. Using y_j as the observed data, the objective function that is minimized is:

$$S = \sum_{j=1}^{n} \left[y_j - \sum_{i=1}^{I} a_i B_{i,q}(x_j) \right]^2 \tag{6.19}$$

where $B_{i,q}(x_j)$ is the value of the B-spline q at x_j, $\sum_{i=1}^{I} a_i B_{i,q}(x_j)$ is the sum of B-splines (Price, 2018).

P-spline is abbreviation for penalized B-spline and refers to using the B-spline representation where the coefficients are established partly by the data that will be fitted, and partially by an additional penalty function that targets to force smoothness to prevent overfitting. P-splines have been established by (Eilers & Marx, 1996) building up on B-splines. "The B-spline basis functions are strictly local, so each basis function is only non-zero over the intervals between m+3 adjacent knots, where m+1 is the order of the basis" (Wood, 2017). A spline of order m + 1 is written as

$$S(x) = \sum_{i=1}^{k} \beta_i{}^m(x)\beta_i \tag{6.20}$$

where,

$$\beta_i{}^m(x) = \frac{x-x_i}{x_{i+m+1}-x_i}\beta_i{}^{m-1}(x) + \frac{x_{i+m+2}-x}{x_{i+m+2}-x_{i+1}}\beta_{i+1}{}^{m-1}(x) \tag{6.21}$$

where, i=1, …, k and

$$\beta_i{}^{-1}(x) = \begin{cases} 1 & x_i \le x \le x_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{6.22}$$

The penalty function is described as the squared difference between adjacent β_i values.

$$P = \sum_{i=1}^{k-1}(\beta_{i+1} - b_i)^2 \tag{6.23}$$

P-splines are simple to set up and enables a decent amount of flexibility by the ability to combine any order of penalty with any order of B-spline basis. There are others spline such as cubic regression spline , cyclic cubic regression spline, thin plate regression spline (Claeskens et al., 2009; Eilers & Marx, 2010; Welham et al., 2007).

6.7 P-IRLS (Penalized Iteratively Re-Weighted Least Squares)

The estimation of generalized additive model involves the estimation of the smoothing parameters and obtaining the model coefficients of the maximum penalized likelihood function. The choice of the basis function and the smoothing parameter is central in GAM estimation (Marra & Wood, 2012). The smooth terms are now represented as a linear combination of the basis function, b_{jk}, and the unknown regression parameters, β_{jk}:

$$f_j(x_j) = \sum_{k=1}^{qj} \beta_{jk} b_{jk}(x_j) \tag{6.24}$$

where x_j can be a vector and the b_{jk} are the coefficients of the smooth to be estimated. Substituting each smooth term $f_j(x_j)$, by their bases then the GAM can now be written as:

$$g(\mu_j) = \eta_j = \boldsymbol{X_j\beta} \tag{6.25}$$

where $\boldsymbol{X_j} = b_{ji}(x_{jk})$. We can compose the likelihood $l_p\beta$ to estimate the model. GAMs are approximated by penalized likelihood maximization because estimating by ordinary maximum likelihood can cause overfitting (Wood, 2006). The penalized likelihood function can be stated as:

$$l_p(\beta) = l(\beta) - \frac{1}{2}\sum_{j=1}^{p} \lambda_j \, l(\beta) - \frac{1}{2}\sum_{j=1}^{p} \lambda_j \, \beta' S_j \, \beta$$

$$= \; l(\beta) - \tfrac{1}{2} \; \beta' S \, \beta \tag{6.26}$$

where $S = \sum_{j=1}^{p} \lambda_j S_j$ and λ_j are the smoothing parameters assumed to be known, controlling the tradeoff between goodness of fit of the model smoothness (Antoniadis et al., 2011). To maximize l_p we set derivatives with respect to β_j to zero:

$$\frac{\partial l_p}{\partial \beta_j} = \frac{\partial l}{\partial \beta_j} - [S\beta]_j = \frac{1}{\phi}\sum_{i=1}^{n} \frac{y_i - \mu_i}{V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_j} - [S\beta]_j = 0 \tag{6.27}$$

where, $[.]_j$ denotes the j^{th} row of a vector. These equations are the same as those that would be solved to maximize the penalized nonlinear least squares problem:

$$S_p = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{Var(Y_i)} + \; \beta' S \, \beta \tag{6.28}$$

where, μ_i depends non-linearly on β and assuming $Var(Y_i)$ terms are known. To solve these equations, the iterative method is required. It can be shown that in the proximity of a certain coefficient vector estimate $\hat{\beta}^{[k]}$.

$$S_p \simeq \left\| \sqrt{\omega^{[k]}} \left(z^{[k]} - x\beta \right) \right\|^2 + \; \beta' S \, \beta \tag{6.29}$$

Given a model's link function g, $z^{[k]}$ is a vector of pseudo-data and $\omega^{[k]}$ is a diagonal matrix, W has diagonal elements ${\omega_i}^{[k]} = \frac{1}{V({\mu_i}^{[k]}) g'({\mu_i}^{[k]})^2}$ and the pseudo-data is defined as

$${z_i}^{[k]} = g'(\mu^{[k]})(y_i - {\mu_i}^{[k]}) + \; X_i \, \hat{\beta}^{[k]} \tag{6.30}$$

where $z^k$is a vector of pseudo-data with elements $z_i{}^{[k]}$ and g is the model link function. Taking the assumption that smoothing parameters are known, the maximum penalized likelihood estimates, $\hat{\beta}$, can be estimated by repeating the two steps below to convergence:

Step 1: Use current β^k, calculate the pseudo-data $z^{[k]}$ and $\omega_i{}^{[k]}$.

Step 2: Minimize equation 6.29 with respect to β, to find $\hat{\beta}^{[k+1]}$. Calculate the linear predictor $\eta^{[k+1]} = X\beta^{[k+1]}$ and fitted values $\mu_i{}^{[k+1]} = g^{-1}(\eta_i{}^{[k-1]})$. After each iteration, we obtain new values of the coefficients μ and β; and update the weights ω_i and pseudo data z_i. This Iteration is repeated until convergence (Wood, 2006).

6.8 Generalized Additive Logistic Regression Model and Application

The GAMs substitute the linear predictor with an additive predictor. GAMs doesn't take into consideration the assumption of linearity and allows you to reveal relationships between the independent and dependent variable that might be overlooked (Brockett et al., 2014). The GAM logistic model is given by

$$\pi(x) = \frac{exp\left(S_0+\sum_{j=1}^{p} S_j(x_{ij})\right)}{1+exp\left(S_0+\sum_{j=1}^{p} S_j(x_{ij})\right)} \tag{6.31}$$

The GAM is used to fit the DHS dataset to determine the factors associated with child mortality in Lesotho. The previous chapters applied the parametric methods of logistic regression, survey logistic regression and the GLMM. GAMs combine both features of parametric and non-parametric regression hence the term, semi-parametric regression (Liu, 2008). The semi-parametric logistic model is written as

$$logit(\pi(x)) = \beta_0 + \sum_{j=1}^{p} \beta_j(x_{ij}) + \sum_{j=p+1}^{q} S_j(x_{ij}) \tag{6.32}$$

6.9 Fitting the GAM Model using the GAM PROC

"The generalized additive models fitted by the GAM procedure combine an assumption of additivity" (Stone, 1985) that allows a lot of nonparametric relationships to be considered at the same time and the distributional flexibility of GLMs (Nelder & Wedderburn, 1972).

PROC GAM gives a lot of flexibility in modeling the predictor and response relationships. For predictors in regression models, the additive models assume nonparametric smoothing splines (Cai, 2008). "Generalized linear models make the assumption that dependent variable depends on additive predictors through a monotonic nonlinear link function specified by a distribution member in the exponential family" (Hardin et al., 2007). The combination of the two assumptions allows GAMs to be used in a lot of modeling scenarios (Dunteman & Ho, 2005).

6.9.1 Fitting the Model

The PROC GAM procedure was used to fit the model, with flexible spline terms for each of the predictors. Three degrees of freedom are chosen for our model as this is the default corresponding to a smoothing spline with the complexity of a cubic polynomial. Two of these degrees of freedom are for the nonlinear spline portion and one degree of freedom is taken up by the linear part of the fit.

Table 6.1 illustrates the parameter estimates and the linear portion for parametric part of the model, t-values, standard errors and p-values. The second part of the table shows smoothing parameters, degrees of freedom, GCV value for each predictor and the number of unique observations. Currently breastfeeding is negatively associated with child mortality with a p-value equals 0.0041. The type of residence is also negatively associated with child mortality at 5% significant level with a p-value equals 0.0010. The predictor variables child-birth order and number of children living are observed significantly associated with child mortality since the corresponding p-value is below 0.05 (p-value=<.0001). The predictor mother's age in linear portion is similarly significantly associated with child mortality with a p-value equals 0.0003.

Table 6.1: Analytical information about the fitted model(2009)

Regression Model Analysis

Parameter Estimates

Parameter	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	-3.0200	0.3427	-8.81	<.0001
Currently Breastfeeding	-0.42250	0.14701	-2.87	0.0041
Urban	-0.6327	0.1929	-3.28	0.0010
Linear(Birth order number)	0.13937	0.08695	16.03	<.0001
Linear(Mother's age)	0.05360	0.01465	3.66	0.0003
Linear(Number of children living)	-1.80302	0.10123	-17.81	<.0001

Smoothing Model Analysis

Fit Summary for Smoothing Components

Component	Smoothing Parameter	DF	GCV	Num Unique Obs
Spline(Birth order number)	0.971370	3	3.095261	13
Spline(Mother's age)	0.999214	3	1.356085	35
Spline(Number of children living)	0.992922	3	43.571982	13

In the 2014 dataset; Currently breastfeeding is negatively associated with child mortality with a p-value<0.0001. The type of residence is also negatively associated with child mortality at 5% significant level (p-value=0.0377). The predictor variable child-birth order and number of children living was observed significantly associated with child mortality since the corresponding p-value is below 0.05. The predictor mother's age in linear portion was not significant (p-value=0.3285). This could be the result of some part of significance being taken by non-linear part.

Table 6.2: Analytical information about the fitted model(2014)

Regression Model Analysis

Parameter Estimates

Parameter	Parameter Estimate	Standard Error	t Value	Pr > |t|

Intercept	-2.53668	0.43272	-5.86	<.0001
Currently Breastfeeding	-0.95175	0.21268	-4.48	<.0001
Urban	-0.42264	0.20334	-2.08	0.0377
Linear(Birth order number)	1.45516	0.11146	13.06	<.0001
Linear(Mother's age)	0.01796	0.01838	0.98	0.3285
Linear(Number of children living)	-1.75498	0.1258	-13.95	<.0001

Smoothing Model Analysis
Fit Summary for Smoothing Components

Component	Smoothing Parameter	DF	GCV	Num Unique Obs
Spline(Birth order number)	0.950529	3	2.347586	12
Spline(Mother's age)	0.999213	3	4.114624	35
Spline(Number of children living)	0.946329	3	7.302263	12

The amount of smoothing is indicated by the degree of freedom. An increase in smoothing means less degree of freedom which means a higher span. In the smoothing model analysis, the smoothing parameter for all components was close to one in both datasets(table 6.1 and 6.2) and the degree of freedom that corresponds is 3. The Analysis of Deviance Table 6.3. gives a chi square-test from making a comparison of the deviance between the full model and the model without the nonparametric component of this variable for each smoothing effect in the model. This table is the critical part of the GAM results (Table 6.3 & Table 6.4). The analysis of deviance results shows that the nonparametric effect of Birth order number and number of children living is significant while the mother's age is insignificant at the 5% level.

Table 6.3: Analysis of Deviance

Smoothing Model Analysis
Analysis of Deviance

Source	DF	Sum of Squares	Chi-Square	Pr > ChiSq
Spline(Birth order number)	3	22.440591	22.4406	<.0001
Spline(Mother's age)	3	4.102809	4.1028	0.2506
Spline(Number of children living)	3	131.078792	131.0788	<.0001

The analysis of deviance results for the 2014 dataset indicate that the nonparametric effect of birth order number and number of children living is significant with p-value below 0.05 and the nonparametric effect of mother's age is not significant at the 5% level.

Table 6.4: Analysis of Deviance

Smoothing Model Analysis Analysis of Deviance				
Source	**DF**	**Sum of Squares**	**Chi-Square**	**Pr > ChiSq**
Spline(Birth order number)	3	26.14147	26.1415	<.0001
Spline(Mother's age)	3	4.454243	4.4542	0.2164
Spline(Number of children living)	3	113.3362	113.3362	<.0001

The plots below (Figure 6.4 and Figure 6.5) have the estimated smoothing spline function with the linear effect subtracted out. B5 is the response variable of child survival status. The plot includes a 95% confidence band for the whole curve. We visually examine where this band does not include zero to get indication of where significant nonlinearity takes place.

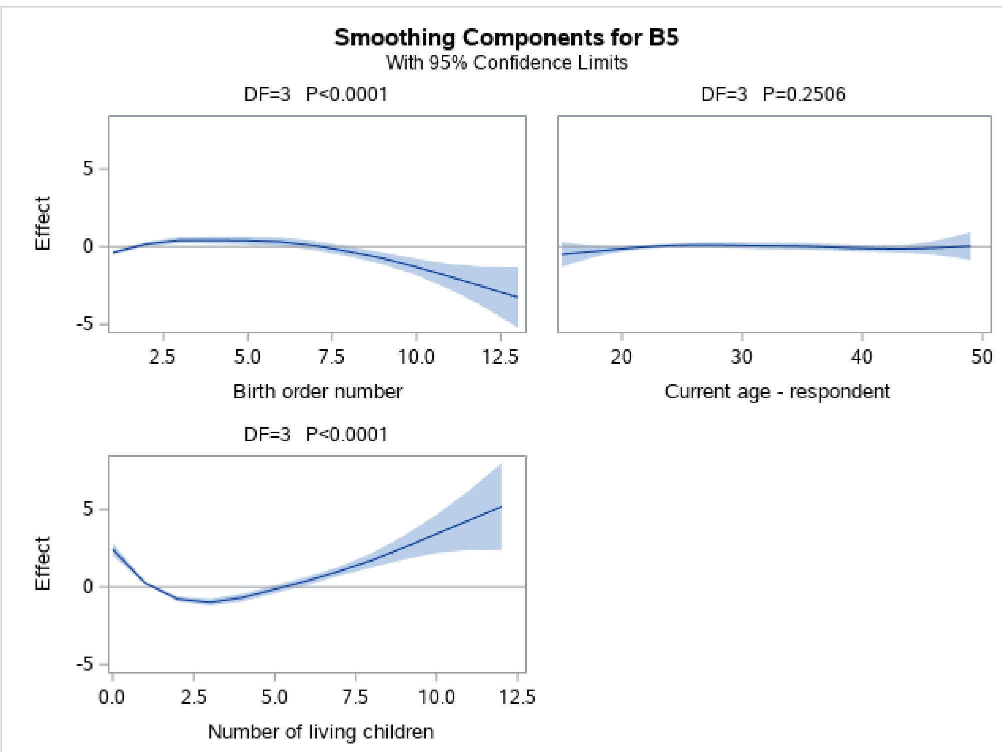

Figure 6.4: Partial Prediction for Each Predictor (2009).

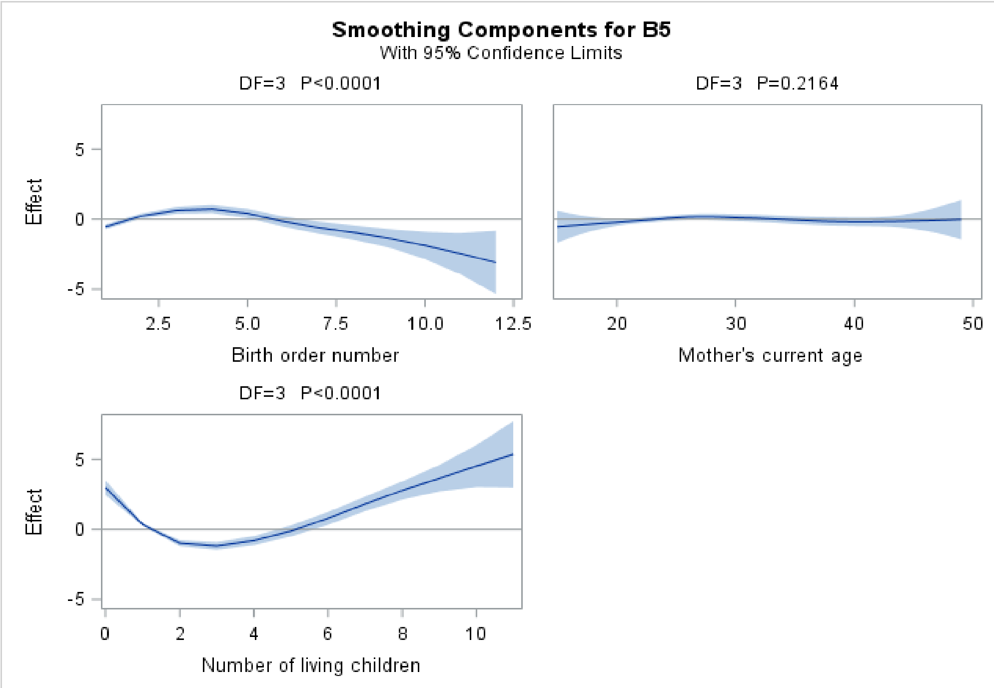

Figure 6.5: Partial Prediction for Each Predictor (2014).

Figure 6.4 illustrates the partial predictions that correspond to the number of children living and the birth order number have a quadratic pattern and mother's age doesn't have a quadratic pattern. This means the under-five child mortality was associated with a quadratic pattern for number of children living and birth order number. The 95% confidence limits for the mother's age includes the zero axis, also in agreement that the term is insignificance.

For the 2014 dataset; The plots display that the partial predictions that correspond to the number of children living, and the birth order number have a quadratic pattern while the mother's age doesn't have a quadratic pattern. This means that child mortality was associated with a quadratic pattern for number of children living and the birth order number. The 95% confidence limits for the mother's age includes the zero axis, also in agreement that the term is insignificance. There is no difference between the results for 2009 data and 2014 data.

Chapter 7

7 Discussion

7.1 Conclusions and Recommendations

This research was executed with the purpose to determine which factors are related to the child mortality in Lesotho. The under-five child survival status was researched applying statistical models (LR, SLR, GLMMs, and GAMs) using the 2009 and the 2014 Lesotho Demographic Health Surveys. The 2009/2014 LDHS was executed by the Ministry of Health and Social Welfare (MOHSW) having the assistance of the Bureau of Statistics (BOS). The 2009 and the 2014 LDHS were designed to give estimates of demographic and health indicators for Lesotho, in the urban and rural parts, and for individual regions (Thaba Tseka, Maseru, Qacha's Nek, Butha-Buthe, Quthing, Mokhotlong, Berea, Mohale's Hoek, Leribe and Mafeteng).

We have looked at the effect of specifically economic, social and demographic characteristics of mothers on the probability of under-five mortality in Lesotho. Results showed that several factors influence the chance of under-five mortality. When it comes to decision making, this will be used as a guide to help policy makers in their efforts. It will assist in improving the health of children and the prevention of children's deaths as well as growth. This will speed up a better life being provided to the people and assessing MDG4 goals achievement progress. Therefore, policy makers should strengthen the interventions for child health in order to decrease dying children below age 5. Generalized linear models, Survey logistic regression models, generalized linear mixed models, and generalized additive models were used to identify the risk factors.

In chapter 4, the logistic regression model was used, and it presumed survey data was acquired through SRS and then the survey logistic regression model which takes into consideration how complex the survey design was. The demographic and socio-economic factors were used as predictor variables. The two-way interaction effects were incorporated in the modeling procedure. The results agree with the hypothesis that demographic and socioeconomic factors are significant in affecting child mortality. With logistic regression, the factors age of child, birth order number, wealth index , education attainment, mother's age, weight of child, type of place of residence, number of children living and two-way interaction effect currently breastfeeding by age of child have been established to be significantly associated with child mortality under 5 using data from both 2009 and 2014. In the application of the 2009 data set, wealth index was not significantly associated with under 5 mortality and in the 2014 data set, weight of child was not associated with under five mortality. In both data sets, sex of child, marital status, mothers work status and region were not significantly associated with the under 5 mortality in both data sets. The Hosmer and Lemeshow test was used for model checking and testing the goodness of fit of the logistic regression model. The test indicates that the logistic model is a good fit to the data. The Model checking and goodness of fit using Hosmer-Lemeshow test failed to reject the selected model.

The model was refitted through the survey logistic regression model and generalized linear mixed models. Both models seem to be the good alternative since they account for the

The SLR in chapter 4 and GLMMs in chapter 5 accounts for the complexity of the survey designs and findings in this research illustrates that the models are a good substitute. For the 2009 data, the risk of a child dying who is over 1-year old is greater than that of a child who is below 1 year. The risk of dying for a child who's not breastfed is larger than that of a child who's breastfed. The risk of dying for a child who's from a poor family is higher in comparison to that of a child who's from a rich family. The risk of dying for a child whose mother has no education or up to secondary education is larger than that of a child whose mother has higher education. The risk of dying for a child who's from a rural area is higher compared to that of a child who's from an urban area. The risk of dying for a child with birth order number is 2-4 or >4 is higher in comparison to that of a child who's first born. The risk of dying for a child whose household has 2-4 or >4 number of children living is larger than that of a child whose household has <2 number of children living. This could be because households with less children allows the mother to focus on the child and financially makes the household better. For the 2014 data, the results are similar. The risk of dying for a child whose weight is small is higher in comparison to that of a child whose weight is large. Design effects were used to compare parameter estimates from both models. Most of the DEFF values are over one meaning that standard errors for LR model are smaller compared to standard errors for survey. This implies that there was a variance under-estimation when applying the logistic regression, that presumes that the data were sampled by means of SRS.

In chapter 6, the semiparametric regression approach is applied; the under-five mortality for both data sets was observed to be significantly associated with the quadratic pattern of child-birth order number and number of children living. The under-five child mortality for both data sets is not associated with mother's age as there is no quadratic effect at 5% significant level. In general, this research shows that age of child, birth order number, weight of child at birth, breastfeeding, wealth index , education attainment, mother's age, type of place of residence, number of children living were found to be determining factors of the child mortality in Lesotho.

The discovered factors may be used as guidance to policy makers on increasing the speed of providing an improved health service. Policy implications since Lesotho has devoted to the MDGs, the number 4 goal which is decreasing child mortality, the country should be persistent in its attempts to achieve this goal. The research paper has illustrated key policy implications that the government needs to give attention to. The findings in this paper suggests that the child health status can improve in Lesotho. There needs to be more focus from the policy makers on the significant factors in order to develop more strategies that can improve child wellbeing by reducing child mortality in Lesotho.

Mothers' education has an impact on child mortality under-five and women empowerment and enablement through education should be encouraged. Children whose mothers have completed high school have a reduced chance of child mortality by nearly 55%. We cannot disregard the education of the mother as most studies have revealed that it has influence on child mortality (Caldwell & McDonald, 1982; Hobcraft, 1993; Kiross et al., 2019). The effect of mother's education is important and that is because mothers with education are more likely to get pre-birth care and are informed about getting health care. Education has impact because women are more informed and hence less unexpected pregnancies (Pradhan, 2015). Breastfeeding should be encouraged to enhance the health of under five children and hence reduce the risk of child mortality (Gilbert & Gichuhi, 2014). Breastfeeding throughout the first six months of the child being born and persistence of breastfeeding while adding other foods until the child is two are suggested (Arifeen et al., 2001; Shifa et al., 2018). Another way to improve child mortality is by controlling the number of children a mother gives birth to, because the number of children living is also a contributing factor. Enhancing the mothers' economic status will improve the child's basic needs and therefore can result in the decline of child mortality.

The place of residence has significant influence on child mortality. The odds of a child death were lesser in the urban areas than in rural areas. This is expected as the level of development is higher for urban than rural areas. Intervention programs could help with rural areas having improve sources of drinking water. Advancing the work status of the mothers can help the mother's economic status, therefore have improvement in the basic needs of their children. The government should continually observe and assess current programs to be able to review and develop new programs that are more applicable to the factors that influence child survival status.

7.2 Limitations

The LDHS data depends on the mothers reporting about the survival and conditions of the children they have birthed and that impacts the death and the birth dates accuracy, especially when no paperwork was verified. Deaths that could have occurred after the survey was taken will not be considered. Deaths reported in this research may not be a portion of the true under five child mortality in Lesotho. Furthermore, the place of residence of certain respondents may be different over the period from the date that a child was born. More work can be done on this study, such as focusing on GAM to include random effects as well as account for the missing values.

Appendix A

Codes Used to fit the models

Variable used to fit the models:

B8- Current age of child, V101- Region, V102- Type of place of residence, V190A- Wealth index, V404- Currently breastfeeding, V714- Mothers currently working, V149A- Education attainment, V501AA- Marital status, M18A- Weight of the child at birth, BORDA- Birth order number, V218A- Number of children living, B4- Sex of child and V012- Mother's age.

A.1 SAS code for 2009 LDHS: Logistic Regression

```
PROC IMPORT OUT= WORK.LDHS2009
DATAFILE= "C:\Users\211534620\Documents\2014 fin\LSKR61FL
          original.SAV"
DBMS=SPSS REPLACE;

RUN;

Proc logistic Data=WORK.LDHS2009 plots=all;
Class B8(ref="0") B4(ref="Male") V101(ref=" Thaba-Tseka")
      V102(ref="Urban") V190A(ref="Rich") V404(ref="Yes")
      714(ref="Employed")V149A(ref=" Higher Education")
      V501AA(ref="Not Married") M18A(ref="Large") BORDA(ref=" first
      birth") V218A(ref="<2 children")/ param=glm;
Model B5(event="No")=B8 V012 B4 M18A V404  V501AA V190A V149A V714
      V102 V101 BORDA V218A B8*v404/lackfit;
Run;
```

A.2 SAS code for 2009 LDHS: Survey Logistic Regression

```
Proc SurveyLogistic Data=WORK.LDHS2009;
stratum V023;
cluster V021;
weight V005A;
Class B8(ref="0") B4(ref="Male") V101(ref=" Thaba-Tseka")
      V102(ref="Urban")V190A(ref="Rich") V404(ref="No")
      V714(ref="Unemployed")V149A(ref=" No Education")
      V501AA(ref="Not Married") M18A(ref="Large")BORDA(ref="first
      birth")  V218A(ref="<2 children")/ param=glm;
Model B5(event="No")=B8 V012 B4 M18A V404 V501AA V190A V149A V714 V102
      V101 V218A BORDA B8*V404;
Run;
```

A.3 SAS code for 2009 LDHS: Generalized Linear Mixed Model

```
PROC IMPORT OUT= WORK.LDHS2009
             DATAFILE= "/folders/myfolders/LSKR61FL original.SAV"
            DBMS=SPSS REPLACE;
RUN;

Proc glimmix data=WORK.LDHS2009 method=laplace;
Class V021 B8(ref="0") B4(ref="Male") V101(ref="Thaba-Tseka")
      V102(ref="Urban") V714(ref="Employed")V149A(ref="Higher
Education")
      V190A(ref= "Rich" )V404(ref= "Yes" )V501AA(ref= "Not Married" )
      BORDA(ref= "first birth" )  V218A(ref="<2 children")
      M18A(ref="Large");
Model B5(event="No")=B8 V012 B4 M18A V404 V501AA V190A V149A V714 V102
      V101 V218A BORDA B8*V404/ link=logit DIST=binary oddsratio
      solution;
lsmeans B4 M18A V501AA V190A V149A V714 V102 V101 V218A BORDA B8*V404/
       plot=diffplot adjust=tukey alpha=0.05;
lsmeans B4 M18A V501AA V190A V149A V714 V102 V101 V218A BORDA B8*V404/
       plot=anomplot adjust=nelson alpha=0.05;
random intercept /subject=V021 TYPE=VC;
Covtest zerog;
run;
```

A.4 SAS code for 2009 LDHS: Generalized Additive Model

```
Ods graphics on; ods html;
Proc gam data=LDHS2009 desc
Plots=components(commonaxes clm);
Class V404 V102;
Model B5(event="No")= param(V404) param(v102) spline(BORD)
      spline(V012) spline(V218)/ dist=binomial link=logit;
run; Ods html close;  ods graphics off
```

A.5 SAS code of 2014 LDHS: Logistic Regression

```
Proc logistic data=WORK.LDHS2014 plots=all;
Class B8(ref="0") B4(ref="Male") V101(ref=" Thaba Tseka")
      V102(ref="Urban") V190A(ref="Rich") V404(ref="Yes")
      V714(ref="Employed")V149A(ref=" Higher Education")
      V501AA(ref="Not Married") M18A(ref="Large") BORDA(ref=" first
      birth") V218A(ref="<2 children")/ param=glm;
Model B5(event="No")=B8 V012 B4 M18A V404 V218A V501AA V190A V149A
      V714 V102 V101  BORDA B8*v404/lackfit;
Run;
```

A.6 SAS code of 2014 LDHS: Survey Logistic Regression

```
Proc surveylogistic data=WORK.LDHS2014;
stratum V023;
cluster V021;
weight V005A;
Class B8(ref="0") B4(ref="Male") V101(ref=" Thaba Tseka")
      V102(ref="Urban") V190A(ref="Rich") V404(ref="No")
      V714(ref="Employed")V149A(ref=" Higher Education")
      V501AA(ref="Not Married") M18A(ref="Large") BORDA(ref=" first
      birth")  V218A(ref="<2 children")/ param=glm;
Model B5(event="No")=B8 V012 B4 v404 M4A V501AA V190A V218A V149A V714
      V102 V101  BORDA B8*V404;
Run;
```

A.7 SAS code of 2014 LDHS: Generalized Linear Mixed Model

```
PROC IMPORT OUT= WORK.LDHS2014
            DATAFILE= "/folders/myfolders/LSKR71FL.SAV"
            DBMS=SPSS REPLACE;
RUN;

Proc glimmix data=WORK.LDHS2014 method=laplace;
Class V021 B8(ref="0") B4(ref="Male") V101(ref="Thaba Tseka")
      V102(ref="Urban") V714(ref="Employed")V149A(ref="Higher Education")
      V190A(ref= "Rich" )V404(ref= "Yes" )V501AA(ref= "Not Married" )
      BORDA(ref= "first birth" )  V218A(ref="<2 children") M18A(ref="Large");
Model B5(event="No")=B8 V012 B4 M18A V404 V501AA V190A V149A V714 V102 V101
      V218A BORDA B8*V404/ link=logit DIST=binary oddsratio solution;
lsmeans B4 M18A V501AA V190A V149A V714 V102 V101 V218A BORDA B8*V404/
plot=diffplot adjust=tukey alpha=0.05;
lsmeans B4 M18A V501AA V190A V149A V714 V102 V101 V218A BORDA B8*V404/
plot=anomplot adjust=nelson alpha=0.05;
random intercept /subject=V021 TYPE=VC;
Covtest zerog;
run;
```

A.8 SAS code of 2014 LDHS: Generalized Additive Model

```
Ods graphics on; ods html;
proc gam data=LDHS2014 desc
Plots= components(commonaxes clm);
Class V404 V102;
Model B5(event="No")= param(V404) param(V102) spline(BORD)
      spline(V012) Spline(V218)/ dist=binomial link=logit;
run; ods html close;

     Ods graphics off;
```

References

Agbozo, F., Abubakari, A., Der, J., & Jahn, A.J.M. (2016). Prevalence of low birth weight, macrosomia and stillbirth and their relationship to associated maternal risk factors in Hohoe Municipality, Ghana. *40*, 200-206.

Ahmad, O.B., Lopez, A.D., & Inoue, M. (2000). The decline in child mortality: a reappraisal. *Bulletin of the World Health Organization, 78*, 1175-1191.

Ahmad, O.B., Lopez, A.D., & Inoue, M.J.B.o.t.W.H.O. (2000). The decline in child mortality: a reappraisal. *78*, 1175-1191.

Ahn, H., & Powell, J.L. (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics, 58*(1-2), 3-29.

Akin, J.S., Birdsall, N., & De Ferranti, D.M. (1987). *Financing health services in developing countries: an agenda for reform* (Vol. 34): World Bank Publications.

Al-Rawi, S.N., Al-Heety, F.A., & Hasan, S.S. (2010). A New Computational Method for Optimal Control Problem with B-spline Polynomials. *Engineering Technology Journal, 28*(18), 5711-5718.

Albert, A., & Anderson, J.A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika, 71*(1), 1-10.

Aldrich, J.H., Nelson, F.D., & Adler, E.S. (1984). *Linear probability, logit, and probit models*: Sage.

Allison, P.D. (2012). *Logistic regression using SAS: Theory and application*: SAS Institute.

Allison, P.D. (2014). *Measures of fit for logistic regression.* Paper presented at the Proceedings of the SAS Global Forum 2014 Conference.

Altman, D., Machin, D., Bryant, T., & Gardner, M. (2013). *Statistics with confidence: confidence intervals and statistical guidelines*: John Wiley & Sons.

Alvarez Meza, A.M., Daza Santacoloma, G., Acosta Medina, C.D., & Castellanos Dominguez, G. (2012). Parameter selection in least squares-support vector machines regression oriented, using generalized cross-validation. *Dyna, 79*(171), 23-30.

Animaw, W., Taye, W., Merdekios, B., Tilahun, M., & Ayele, G. (2014). Expanded program of immunization coverage and associated factors among children age 12–23 months in Arba Minch town and Zuria District, Southern Ethiopia, 2013. *BMC public health, 14*(1), 464.

Antoniadis, A., Gijbels, I., & Nikolova, M. (2011). Penalized likelihood regression for generalized linear models with non-quadratic penalties. *Annals of the Institute of Statistical Mathematics, 63*(3), 585-615.

Archer, K.J., & Lemeshow, S. (2006). Goodness-of-fit test for a logistic regression model fitted using survey sample data. *The Stata Journal, 6*(1), 97-105.

Archer, K.J., Lemeshow, S., & Hosmer, D.W. (2007). Goodness-of-fit tests for logistic regression models when data are collected using a complex sampling design. *Computational Statistics & Data Analysis, 51*(9), 4450-4464.

Arifeen, S., Black, R.E., Antelman, G., Baqui, A., Caulfield, L., & Becker, S. (2001). Exclusive breastfeeding reduces acute respiratory infection and diarrhea deaths among infants in Dhaka slums. *Pediatrics, 108*(4), e67-e67.

Arnold, T.W. (2010). Uninformative parameters and model selection using Akaike's Information Criterion. *The Journal of Wildlife Management, 74*(6), 1175-1178.

Asparouhov, T., & Muthen, B. (2006). *Multilevel modeling of complex survey data.* Paper presented at the Proceedings of the joint statistical meeting in seattle.

Austin, P.C., & Steyerberg, E.W. (2012). Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. *BMC medical research methodology, 12*(1), 82.

Ayele, D.G., Zewotir, T.T., & Mwambi, H.G. (2012). Prevalence and risk factors of malaria in Ethiopia. *Malaria journal, 11*(1), 195.

Bank, W. (2019). Lesotho Poverty Assessment: Progress and Challenges in Reducing Poverty. In: World Bank.

Bennett, S., Woods, T., Liyanage, W.M., & Smith, D.L. (1991). A simplified general method for cluster-sample surveys of health in developing countries.

Berndt, E.R., & Savin, N.E. (1977). Conflict among criteria for testing hypotheses in the multivariate linear regression model. *Econometrica: Journal of the Econometric Society*, 1263-1277.

Bewick, V., Cheek, L., & Ball, J. (2005). Statistics review 14: Logistic regression. *Critical care, 9*(1), 112.

Bieler, G.S., Brown, G.G., Williams, R.L., & Brogan, D.J. (2010). Estimating model-adjusted risks, risk differences, and risk ratios from complex survey data. *American journal of epidemiology, 171*(5), 618-623.

Binder, H., & Tutz, G. (2008). A comparison of methods for the fitting of generalized additive models. *Statistics Computing, 18*(1), 87-99.

Black, R.E., Morris, S.S., & Bryce, J. (2003). Where and why are 10 million children dying every year? *The lancet, 361*(9376), 2226-2234.

Bland, J.M., & Altman, D.G. (2000). The odds ratio. *Bmj, 320*(7247), 1468.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., & White, J.-S.S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology evolution, 24*(3), 127-135.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., & White, J.-S.S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution, 24*(3), 127-135.

Bonnefoix, T., Bonnefoix, P., Verdiel, P., & Sotto, J.-J. (1996). Fitting limiting dilution experiments with generalized linear models results in a test of the single-hit Poisson assumption. *Journal of immunological methods, 194*(2), 113-119.

Breslow, N. (2004). *Whither PQL?* Paper presented at the Proceedings of the Second Seattle Symposium in Biostatistics.

Breslow, N.E., & Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88*(421), 9-25.

Brockett, P.L., Chuang, S.-L., & Pitaktong, U. (2014). Generalized additive models and nonparametric regression. *Predictive modeling applications in actuarial science. Predict Model Tech, 1*, 367.

Bruin, J. (2006). Newtest: command to compute new test. *UCLA: Statistical Consulting Group*.

Buja, A., Hastie, T., & Tibshirani, R. (1989). Linear smoothers and additive models. *the Annals of Statistics*, 453-510.

Burger, A., & Silima, T. (2006). Sampling and sampling design. *Journal of Public Administration, 41*(3), 656-668.

Burman, P., & Chaudhuri, P. (2012). On a hybrid approach to parametric and nonparametric regression. In *Nonparametric Statistical Methods and Related Topics: A Festschrift in Honor of Professor PK Bhattacharya on the Occasion of his 80th Birthday* (pp. 233-256): World Scientific.

Bursac, Z., Gauss, C.H., Williams, D.K., & Hosmer, D.W. (2008). Purposeful selection of variables in logistic regression. *Source code for biology and medicine, 3*(1), 17.

Bush, C.A., & MacEachern, S.N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika, 83*(2), 275-285.

Cadeddu, M., Farrokhyar, F., Levis, C., Cornacchi, S., Haines, T., & Thoma, A. (2012). Users' guide to the surgical literature: Understanding confidence intervals. *Canadian Journal of Surgery, 55*(3), 207.

Cai, W. (2008). *Fitting generalized additive models with the GAM procedure in SAS 9.2.* Paper presented at the SAS Global Forum.

Caldwell, J., & McDonald, P. (1982). Influence of maternal education on infant and child mortality: levels and causes. *Health policy education, 2*(3-4), 251-267.
Centor, R.M., & Schwartz, J.S. (1985). An evaluation of methods for estimating the area under the receiver operating characteristic (ROC) curve. *Medical Decision Making, 5*(2), 149-156.
Claeskens, G., Krivobokova, T., & Opsomer, J.D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika, 96*(3), 529-544.
Claeson, M., & Waldman, R.J. (2000). The evolution of child health programmes in developing countries: from targeting diseases to targeting people. *Bulletin of the World Health Organization, 78*, 1234-1245.
Clarke, J.C., & Casey, J. (1995). Social forestry in Lesotho. *Irish Forestry, 52*(1/2), 110-115.
Cleves, M., Gould, W., Gould, W.W., Gutierrez, R., & Marchenko, Y. (2008). *An introduction to survival analysis using Stata*: Stata press.
Cochran, W.G. (1952). The χ2 test of goodness of fit. *The Annals of Mathematical Statistics*, 315-345.
Cochran, W.G. (1977). Simple random sampling. *Sampling Techniques. Third Edition. John Wiley & Sons*.
Cochran, W.G. (2007). *Sampling techniques*: John Wiley & Sons.
Codd, C. (2014). *A Review and Comparison of Models and Estimation Methods for Multivariate Longitudinal Data of Mixed Scale Type.* The Ohio State University,
Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological bulletin, 70*(6p1), 426.
Collett, D. (2002). *Modelling binary data*: Chapman and Hall/CRC.
Consultation, A.E. (2012). The Structural Dete uctural Determinants of Child Well-being.
Coovadia, H.M., Rollins, N.C., Bland, R.M., Little, K., Coutsoudis, A., Bennish, M.L., & Newell, M.-L. (2007). Mother-to-child transmission of HIV-1 infection during exclusive breastfeeding in the first 6 months of life: an intervention cohort study. *The lancet, 369*(9567), 1107-1116.
Cox, D.R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological), 20*(2), 215-232.
Cramer, J.S. (2003). The origins and development of the logit model. *Logit models from economics and other fields, 2003*, 1-19.
Czepiel, S.A. (2002). Maximum likelihood estimation of logistic regression models: theory and implementation. *Available at czep. net/stat/mlelr. pdf*.
De Boor, C., De Boor, C., & Mathématicien, E.-U. (1978). *A practical guide to splines* (Vol. 27): springer-verlag New York.
Ding, C.-S., Haieh, C.-T., Wu, Q., & Pedram, M. (1996). *Stratified random sampling for power estimation.* Paper presented at the Proceedings of International Conference on Computer Aided Design.
Dobson, A.J., & Barnett, A.G. (2008). *An introduction to generalized linear models*: Chapman and Hall/CRC.
Dorfman, D.D., & Alf Jr, E. (1969). Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—Rating-method data. *Journal of mathematical Psychology, 6*(3), 487-496.
Dunteman, G.H., & Ho, M.-H.R. (2005). *An introduction to generalized linear models* (Vol. 145): Sage Publications.
Efron, B. (2004). The estimation of prediction error: covariance penalties and cross-validation. *Journal of the American Statistical Association, 99*(467), 619-632.
Eilers, P.H., & Marx, B.D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 89-102.
Eilers, P.H., & Marx, B.D. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(6), 637-653.

Ettarh, R., Kimani, J.J.R., & Health, R. (2012). Determinants of under-five mortality in rural and urban Kenya. *12*(1).

Eubank, R.L. (1999). *Nonparametric regression and spline smoothing*: CRC press.

Fagerland, M.W., & Hosmer, D.W. (2012). A generalized Hosmer–Lemeshow goodness-of-fit test for multinomial logistic regression models. *The Stata Journal, 12*(3), 447-453.

Fahrmeir, L., & Tutz, G. (2013). *Multivariate statistical modelling based on generalized linear models*: Springer Science & Business Media.

Faraway, J.J. (2016). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*: Chapman and Hall/CRC.

Faturiyele, I., Karletsos, D., Ntene-Sealiete, K., Musekiwa, A., Khabo, M., Mariti, M., Mahasha, P., Xulu, T., & Pisa, P.T. (2018). Access to HIV care and treatment for migrants between Lesotho and South Africa: a mixed methods study. *BMC public health, 18*(1), 668.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters, 27*(8), 861-874.

Fienberg, S.E., & Rinaldo, A. (2007). Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference, 137*(11), 3430-3445.

Fotso, J.-C., Ezeh, A.C., Madise, N.J., & Ciera, J.J.B.P.H. (2007). Progress towards the child mortality millennium development goal in urban sub-Saharan Africa: the dynamics of population growth, immunization, and access to clean water. *7*(1), 1-10.

Friedman, J.H., & Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association, 76*(376), 817-823.

Friedman, J.H., & Stuetzle, W. (1982). *Smoothing of scatterplots*. Retrieved from

Gaffey, M.F., Das, J.K., & Bhutta, Z.A. (2015). *Millennium Development Goals 4 and 5: Past and future progress.* Paper presented at the seminars in fetal and neonatal medicine.

Garenne, M., & Gakusi, E. (2006). Health transitions in sub-Saharan Africa: overview of mortality trends in children under 5 years old (1950-2000). *Bulletin of the World Health Organization, 84*, 470-478.

Gayawan, E., & Ipinyomi, R.A. (2009). A comparison of Akaike, Schwarz and R square criteria for model selection using some fertility models. *Australian Journal of Basic and Applied Sciences, 3*(4), 3524-3530.

Gebretsadik, S., & Gabreyohannes, E. (2016). Determinants of under-five mortality in high mortality regions of Ethiopia: an analysis of the 2011 Ethiopia Demographic and Health Survey data. *International Journal of Population Research, 2016*.

Gellman, B. (2000). AIDS is declared threat to security. *Washington Post, 30*.

Geys1, H., Molenberghs1, G., & Ryan2, L.M. (1997). Pseudo-likelihood inference for clustered binary data. *Communications in Statistics-Theory Methods, 26*(11), 2743-2767.

Gilbert, O., & Gichuhi, W. (2014). Determinants of infant mortality in rural Kenya. *Research on Humanities Social Sciences, 4*(28), 113-121.

Glas, A.S., Lijmer, J.G., Prins, M.H., Bonsel, G.J., & Bossuyt, P.M. (2003). The diagnostic odds ratio: a single indicator of test performance. *Journal of clinical epidemiology, 56*(11), 1129-1135.

Goldstein, H., & Rasbash, J. (1996). Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 159*(3), 505-513.

Golub, G.H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics, 21*(2), 215-223.

Green, P.J. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society: Series B (Methodological), 46*(2), 149-170.

Greg, M. (2017). Gender Profile 2017: Lesotho.

Gueorguieva, R. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling, 1*(3), 177-193.

Guffey, D. (2013). *Hosmer-lemeshow goodness-of-fit test: Translations to the cox proportional hazards model.*

Guisan, A., Edwards Jr, T.C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling, 157*(2-3), 89-100.

Hall, P., & Titterington, D. (1987). Common structure of techniques for choosing smoothing parameters in regression problems. *Journal of the Royal Statistical Society: Series B (Methodological), 49*(2), 184-198.

Handayani, D., Notodiputro, K.A., Sadik, K., & Kurnia, A. (2017). *A comparative study of approximation methods for maximum likelihood estimation in generalized linear mixed models (GLMM).* Paper presented at the AIP Conference Proceedings.

Hardin, J.W., Hardin, J.W., Hilbe, J.M., & Hilbe, J. (2007). *Generalized linear models and extensions*: Stata press.

Härdle, W.K., Müller, M., Sperlich, S., & Werwatz, A. (2012). *Nonparametric and semiparametric models*: Springer Science & Business Media.

Hassan, F.M. (2002). *Lesotho: Development in a Challenging Environment: A Joint World Bank-African Development Bank Evaluation*: The World Bank.

Hastie, T., & Tibshirani, R. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association, 82*(398), 371-386.

Hastie, T.J., & Tibshirani, R.J. (1990). *Generalized additive models* (Vol. 43): CRC press.

Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection–a review and recommendations for the practicing statistician. *Biometrical Journal, 60*(3), 431-449.

Hibberts, M., Johnson, R.B., & Hudson, K. (2012). Common survey sampling techniques. In *Handbook of survey methodology for the social sciences* (pp. 53-74): Springer.

Hilbe, J.M. (1993). Generalized additive models software. In: JSTOR.

Hilbe, J.M. (2009). *Logistic regression models*: Chapman and hall/CRC.

Hilliard, R. (2016). Paediatric challenges in Sub-Saharan Africa. *Journal of Health Specialties, 4*(1), 12.

Hobcraft, J. (1993). Women's education, child welfare and child survival: a review of the evidence. *Health Transition Review*, 159-175.

Hobcraft, J.N., McDonald, J.W., & Rutstein, S.O. (1984). Socio-economic factors in Infant and child mortality: A cross-national comparison. *Population Studies, 38*(2), 193-223. doi:10.1080/00324728.1984.10410286

Honwana, F.E., & Melesse, S.F.J.T.O.P.H.J. (2017). Socio-Economic and demographic determinants of under-five mortality in Ethiopia, 2011. *10*(1).

Hoover, D.R., Rice, J.A., Wu, C.O., & Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika, 85*(4), 809-822.

Horton, N.J., & Fitzmaurice, G.M. (2004). Regression analysis of multiple source and multiple informant data from complex survey samples. *Statistics in medicine, 23*(18), 2911-2933.

Hosmer, D.W., Hosmer, T., Le Cessie, S., & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in medicine, 16*(9), 965-980.

Hosmer, D.W., & Lemeshow, S. (2000). *Applied logistic regression*: Wiley New York.

Hosmer Jr, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). *Applied logistic regression* (Vol. 398): John Wiley & Sons.

Hurvich, C.M., Simonoff, J.S., & Tsai, C.L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Methodological), 60*(2), 271-293.

Hutcheson, G.D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*: Sage.

Institute, S. (2015). *Base SAS 9.4 procedures guide*: SAS Institute.

Jain, N., Gupta, R., & Gupta, R. (2017). Generalized Additive and Generalized Linear Modeling for Children Diseases. *Quest Journals-Journal of Research in Applied Mathematics, 3*(5), 0-10.

Jennrich, R.I., & Robinson, S.M. (1969). A Newton-Raphson algorithm for maximum likelihood factor analysis. *Psychometrika, 34*(1), 111-123.

Jennrich, R.I., & Sampson, P. (1976). Newton-Raphson and related algorithms for maximum likelihood variance component estimation. *Technometrics, 18*(1), 11-17.

Jorgensen, M. (2006). Iteratively reweighted least squares. *Encyclopedia of Environmetrics, 3*.

Kachman, S.D. (2000). *An introduction to generalized linear mixed models.* Paper presented at the Proceedings of a Symposium at the Organizational Meeting for a NCR Coordinating Committee on “Implementation Strategies for National Beef Cattle Evaluation”. Athens.

Kalipeni, E. (2000). Health and disease in southern Africa: a comparative and vulnerability perspective. *J Social Science Medicine*

50(7-8), 965-983.

Kaundjua, M.B. (2013). The determinants of the child mortality rate in rural Namibia.

Khodaee, G.H., Khademi, G., & Saeidi, M. (2015). Under-five Mortality in the World (1900-2015). *International Journal of Pediatrics, 3*(6.1), 1093-1095.

Kiross, G.T., Chojenta, C., Barker, D., Tiruye, T.Y., & Loxton, D. (2019). The effect of maternal education on infant mortality in Ethiopia: A systematic review and meta-analysis. *PloS one, 14*(7), e0220076.

Kish, L. (1965). Sampling organizations and groups of unequal sizes. *American sociological review*, 564-572.

Kish, L., & Frankel, M.R. (1974). Inference from complex samples. *Journal of the Royal Statistical Society: Series B (Methodological), 36*(1), 1-22.

Kuk, A.Y. (1999). Laplace importance sampling for generalized linear mixed models.

Lalonde, T.L., Nguyen, A.Q., Yin, J., Irimata, K., & Wilson, J.R. (2013). Modeling correlated binary outcomes with time-dependent covariates. *Journal of Data Science, 11*(4).

Lani, J. (2010). Assumptions of logistic regression. *Statistics Solutions*.

LDHS. (2014). Ministry of Health [Lesotho] and ICF International. 2016. Lesotho Demographic and Health Survey 2014. Maseru, Lesotho: Ministry of Health and ICF International. .

Lee, Y., & Nelder, J.A. (2006). Double hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics), 55*(2), 139-185.

Lee, Y., Nelder, J.A., & Pawitan, Y. (2018). *Generalized linear models with random effects: unified analysis via H-likelihood* (Vol. 153): CRC Press.

Lemani, C. (2013). *Modelling covariates of infant and child mortality in Malawi.* University of Cape Town,

Leowski, J. (1986). Mortality from acute respiratory infections in children under 5 years of age: global estimates.

Levy, P.S., & Lemeshow, S. (2013). *Sampling of populations: methods and applications*: John Wiley & Sons.

Levy, R. (2012). Probabilistic models in the study of language. *NA, NA*.

Li, W., & Nyholt, D.R. (2001). Marker selection by Akaike information criterion and Bayesian information criterion. *Genetic Epidemiology, 21*(S1), S272-S277.

Liao, T.F. (1994). *Interpreting probability models: Logit, probit, and other generalized linear models*: Sage.

Littell, R.C., Pendergast, J., & Natarajan, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Statistics in medicine, 19*(13), 1793-1819.

Liu, H. (2008). Generalized additive model. *Department of Mathematics Statistics University of Minnesota Duluth: Duluth, MN, USA*.

Liu, L., Hill, K., Oza, S., Hogan, D., Chu, Y., Cousens, S., Mathers, C., Stanton, C., Lawn, J., & Black, R.E. (2016). Levels and causes of mortality under age five years. In.

Liu, L., Oza, S., Hogan, D., Chu, Y., Perin, J., Zhu, J., Lawn, J.E., Cousens, S., Mathers, C., & Black, R.E. (2016). Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *The lancet, 388*(10063), 3027-3035.

Liu, Q., & Pierce, D.A. (1994). A note on Gauss—Hermite quadrature. *Biometrika, 81*(3), 624-629.

Liu, X. (2015). *Methods and applications of longitudinal data analysis*: Elsevier.

Long, J.S., & Freese, J. (2006). *Regression models for categorical dependent variables using Stata*: Stata press.

Lumley, T. (2004). Analysis of complex survey samples. *J Stat Softw, 9*(1), 1-19.

Lumley, T., & Scott, A. (2014). Tests for regression models fitted to survey data. *Australian & New Zealand Journal of Statistics, 56*(1), 1-14.

MacKinnon, J.G. (2015). 1. Wald Tests.

Mahmoud, H.F. (2019). Parametric versus Semi and Nonparametric Regression Models. *arXiv preprint arXiv:.10221*.

Malthus, T.R. (1766). Encyclopedia> Thomas Malthus. *Birth*.

Marcotte, D. (1995). Generalized cross-validation for covariance model selection. *Mathematical geology, 27*(5), 659-672.

Marra, G., & Wood, S.N. (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics, 39*(1), 53-74.

Marx, B.D., & Eilers, P.H. (1998). Direct generalized additive modeling with penalized likelihood. *Computational Statistics Data Analysis, 28*(2), 193-209.

Masangwi, S.J., Ferguson, N., Grimason, A., Morse, T., Zawdie, G., & Kazembe, L. (2010). Household and community variations and nested risk factors for diarrhoea prevalence in southern Malawi: a binary logistic multi-level analysis. *International journal of environmental health research, 20*(2), 141-158.

Matlosa, K. (1999). *Aid, development and democracy in Lesotho, 1966-1996.* Paper presented at the a workshop organised by the Centre for Southern African Studies on “Aid, Development and Democracy in Southern Africa”, University of the Western Cape, Cape Town, South Africa.

McAllister, D.A., Liu, L., Shi, T., Chu, Y., Reed, C., Burrows, J., Adeloye, D., Rudan, I., Black, R.E., & Campbell, H. (2019). Global, regional, and national estimates of pneumonia morbidity and mortality in children younger than 5 years between 2000 and 2015: a systematic analysis. *The Lancet Global Health, 7*(1), e47-e57.

McCullagh, P. (2019). *Generalized linear models*: Routledge.

McCullagh, P., & Nelder, J. (1989). Generalized linear models., 2nd edn.(Chapman and Hall: London). *Standard book on generalized linear models*.

McCulloch, C.E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association, 92*(437), 162-170.

McCulloch, C.E., & Neuhaus, J.M. (2005). Generalized linear mixed models. *Encyclopedia of biostatistics, 4*.

Melesse, S., Sobratee, N., & Workneh, T. (2016). Application of logistic regression statistical technique to evaluate tomato quality subjected to different pre-and post-harvest treatments. *Biological Agriculture & Horticulture, 32*(4), 277-287.

Metz, C.E. (2006). Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. *Journal of the American College of Radiology, 3*(6), 413-422.

Meyer, K. (1989). Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genetics Selection Evolution, 21*(3), 317.

Miranda, A., & Rabe-Hesketh, S.J.T.s.j. (2006). Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. *6*(3), 285-308.

Mitchell, M.N. (2012). *Interpreting and visualizing regression models using Stata* (Vol. 5): Stata Press College Station, TX.

Moeti, A. (2007). *Factors affecting the health status of the people of Lesotho.*

MOHSW. (2010). Ministry of Health and Social Welfare (MOHSW) [Lesotho] and ICF Macro. 2010. Lesotho Demographic and Health Survey 2009. Maseru, Lesotho: MOHSW and ICF Macro. .

Moineddin, R., Matheson, F.I., & Glazier, R.H. (2007). A simulation study of sample size for multilevel logistic regression models. *BMC medical research methodology, 7*(1), 34.

Molenberghs, G., & Verbeke, G. (2005). The generalized linear mixed model (GLMM). *Models for Discrete Longitudinal Data*, 265-280.

Molenberghs, G., & Verbeke, G. (2006). *Models for discrete longitudinal data*: Springer Science & Business Media.

Møller, J., & Waagepetersen, R.P. (2007). Modern statistics for spatial point processes. *Scandinavian Journal of Statistics, 34*(4), 643-684.

Morrow-Howell, N., & Proctor, E. (1993). The use of logistic regression in social work research. *Journal of Social Service Research, 16*(1-2), 87-104.

Mosley, W.H., & Chen, L.C. (1984). An analytical framework for the study of child survival in developing countries. *Population development review, 10*, 25-45.

Motsima, T. (2016). The Risk Factors Associated with Under-Five Mortality in Lesotho Using the 2009 Lesotho Demographic and Health Survey. *International Journal of Medical Health Sciences, 10*(1), 43-51.

Müller, H.-G., & Stadtmüller, U. (2005). Generalized functional linear models. *the Annals of Statistics, 33*(2), 774-805.

Müller, S., Scealy, J.L., & Welsh, A.H. (2013). Model selection in linear mixed models. *Statistical Science, 28*(2), 135-167.

Myung, I.J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology, 47*(1), 90-100.

Nadeem, K. (2013). Estimability and likelihood inference for general hierarchical models using data cloning.

Nannan, N., Dorrington, R., Laubscher, R., Zinyakatira, N., Prinsloo, M., Darikwa, T., & Bradshaw, D. (2010). Under-5 mortality statistics in South Africa. *Cape Town: South African Medical Research Counci*.

Neal, D.J., & Simons, J.S. (2007). Inference in regression models of heavily skewed alcohol use data: A comparison of ordinary least squares, generalized linear models, and bootstrap resampling. *Psychology of Addictive Behaviors, 21*(4), 441.

Nelder, J., & Baker, R.J. (1972). Generalized linear models. Encyclopedia of statistical sciences. In: Wiley, New York.

Nelder, J.A., & Wedderburn, R.W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 135*(3), 370-384.

Omariba, D.W.R., Rajulton, F., & Beaujot, R. (2008). Correlated mortality of siblings in Kenya: The role of state dependence. *Demographic Research, 18*, 311-336.

Onar, A. (2014). Laplace Approximations in B ayesian Lifetime Analysis. *Wiley StatsRef: Statistics Reference Online*.
Organization, W.H., Supply, W.U.J.W., & Programme, S.M. (2015). *Progress on sanitation and drinking water: 2015 update and MDG assessment*: World Health Organization.
Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological modelling, 133*(3), 225-245.
Peng, C.-Y.J., Lee, K.L., & Ingersoll, G.M. (2002). An introduction to logistic regression analysis and reporting. *The journal of educational research, 96*(1), 3-14.
Peterson, B., & Harrell Jr, F.E. (1990). Partial proportional odds models for ordinal response variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 39*(2), 205-217.
Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review/Revue Internationale de Statistique*, 317-337.
Picard, R.R., & Cook, R.D. (1984). Cross-validation of regression models. *Journal of the American Statistical Association, 79*(387), 575-583.
Pollock, D. (1993). Smoothing with cubic splines. In: Queen Mary and Westfield College, Department of Economics.
Population, L. (2019). http://worldpopulationreview.com/countries/lesotho-population/. .07/03/2020
Pourahmadi, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika, 87*(2), 425-435.
Pradhan, E. (2015). Female education and childbearing: A closer look at the data. *World Bank Blog. http://blogs. worldbank. org/health/femaleeducation-and-childbearing-closer-look-data*.
Price, M. (2018). Penalized b-splines and their application with an in depth look at the bivariate tensor product penalized b-spline.
Pritchard, C., & Keen, S.J.S.j.o.p.h. (2016). Child mortality and poverty in three world regions (the west, Asia and sub-Saharan Africa) 1988–2010: evidence of relative intra-regional neglect? *, 44*(8), 734-741.
Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal, 2*(1), 1-21.
Rabe-Hesketh, S., & Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 169*(4), 805-827.
Reed, L. (1929). j. and Berkson, J. *The application of the logistic function to experimental data. Reprint from J. Phys. Chem, 33*, 760-799.
Rice, J., & Rosenblatt, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *The annals of Statistics*, 141-156.
Roberts, G., Rao, N., & Kumar, S. (1987). Logistic regression analysis of sample survey data. *Biometrika, 74*(1), 1-12.
Rust, K.F., & Rao, J. (1996). Variance estimation for complex surveys using replication techniques. *Statistical methods in medical research, 5*(3), 283-310.
Rutstein, S.O. (2000). Factors associated with trends in infant and child mortality in developing countries during the 1990s. *Bulletin of the World Health Organization, 78*, 1256-1270.
Salganik, M.J. (2006). Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *Journal of Urban Health, 83*(1), 98.
Särndal, C.-E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling*: Springer Science & Business Media.
SAS. (2019). https://support.sas.com/documentation/cdl/en/statug/63347/HTML/default/viewer.htm#statug_logistic_sect033.htm.14/10/2019

Sastry, N. (1997). What explains rural-urban differentials in child mortality in Brazil? *Social science medicine, 44*(7), 989-1002.

Satti, H., Motsamai, S., Chetane, P., Marumo, L., Barry, D.J., Riley, J., McLaughlin, M.M., Seung, K.J., & Mukherjee, J.S. (2012). Comprehensive approach to improving maternal health and achieving MDG 5: report from the mountains of Lesotho. *PloS one, 7*(8), e42700.

Schabenberger, O. (2005). Introducing the GLIMMIX procedure for generalized linear mixed models. *SUGI 30 Proceedings, 196*.

Shackman, G. (2001). Sample size and design effect. *Albany Chapter of American Statistical Association*.

Shifa, G.T., Ahmed, A.A., & Yalew, A.W. (2018). Maternal and child characteristics and health practices affecting under-five mortality: A matched case control study in Gamo Gofa Zone, Southern Ethiopia. *PloS one, 13*(8), e0202124.

Shun, Z., & McCullagh, P. (1995). Laplace approximation of high dimensional integrals. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(4), 749-760.

Silverman, B.W. (1984). Spline smoothing: the equivalent variable kernel method. *the Annals of Statistics*, 898-916.

Silverman, B.W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society: Series B (Methodological), 47*(1), 1-21.

Skinner, C., & Wakefield, J. (2017). Introduction to the design and analysis of complex survey data. *Statistical Science, 32*(2), 165-175.

Smyth, G.K. (2003). Pearson's goodness of fit statistic as a score test statistic. *Lecture notes-monograph series*, 115-126.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Methodological), 64*(4), 583-639.

Starkweather, J., & Moske, A.K. (2011). Multinomial logistic regression. *Consulted page at September 10th: http://www. unt. edu/rss/class/Jon/Benchmarks/MLR_JDS_Aug2011. pdf, 29*, 2825-2830.

Steinberg, J. (2005). The Lesotho/Free State Border. *Institute for Security Studies Papers, 2005*(113), 1-1.

Stone, C.J. (1985). Additive regression and other nonparametric models. *the Annals of Statistics*, 689-705.

Sumner, A. (2012). Where do the poor live? *World Development, 40*(5), 865-877.

Sun, L., & Ronnegard, L. (2011). Comparison of different estimation methods for linear mixed models and generalized linear mixed models. *D-level Essay in Statistics, Dalarna University, Sweden, 30p*.

Tan, K.-C. (2014). Why global justice matters. *Journal of Global Ethics, 10*(2), 128-134.

Taylor, P., & Krawchuk, M. (2005). Scale and sensitivity of songbird occurrence to landscape structure in a harvested boreal forest. *Avian Conservation and Ecology, 1*(1).

Tetrault, J.M., Sauler, M., Wells, C.K., & Concato, J. (2008). Reporting of multivariable methods in the medical literature. *Journal of Investigative Medicine, 56*(7), 954-957.

Thomas, S.L., & Heck, R.H. (2001). Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in higher education, 42*(5), 517-540.

Tranmer, M., & Elliot, M. (2008). Binary logistic regression. *Cathie Marsh for census and survey research, paper, 20*.

Tuerlinckx, F., Rijmen, F., Verbeke, G., & De Boeck, P. (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical Statistical Psychology, 59*(2), 225-255.

Tutz, G. (2011). *Regression for categorical data* (Vol. 34): Cambridge University Press.

Uddin, J., & Hossain, Z. (2008). Predictors of infant mortality in a developing country. *Asian Journal of Epidemiology, 1*(1), 1-16.

Unicef. (2017). UNICEF; World Bank. 2017. Lesotho Public Health Sector Expenditure Review. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/29344 License: CC BY 3.0 IGO.

Van Malderen, C., Van Oyen, H., & Speybroeck, N. (2013). Contributing determinants of overall and wealth-related inequality in under-5 mortality in 13 African countries. *Epidemiol Community Health, 67*(8), 667-676.

Venables, W., & Ripley, B. (1999). Generalized linear models. In *Modern Applied Statistics with S-PLUS* (pp. 211-240): Springer.

Vickers, A.J. (2005). Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. *BMC medical research methodology, 5*(1), 35.

Vonesh, E.F. (2012). *Generalized linear and nonlinear models for correlated data: theory and applications using SAS*: SAS Institute.

Vrieze, S.I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological methods, 17*(2), 228.

Wand, M., & Ormerod, J. (2008). On semiparametric regression with O'Sullivan penalized splines. *Australian New Zealand Journal of Statistics, 50*(2), 179-198.

Wang, Y. (2012). Model selection. In *Handbook of computational statistics* (pp. 469-497): Springer.

Wang, Y., & Liu, Q. (2006). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research, 77*(2), 220-225.

Wedderburn, R.W. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss—Newton method. *Biometrika, 61*(3), 439-447.

Wedel, M., & DeSarbo, W.S. (1995). A mixture likelihood approach for generalized linear models. *Journal of Classification, 12*(1), 21-55.

Welham, S.J., Cullis, B.R., Kenward, M.G., & Thompson, R. (2007). A comparison of mixed model splines for curve fitting. *Australian New Zealand Journal of Statistics, 49*(1), 1-23.

WHO, U.J.W.H.O., UNICEF, UNFPA,, & Bank, T.W. (2012). UNFPA, The World Bank. Trends in maternal mortality: 1990 to 2010.

Wilber, S.T., & Fu, R. (2010). Risk ratios and odds ratios for common events in cross-sectional and cohort studies. *Academic Emergency Medicine, 17*(6), 649-651.

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology, 40*(1), 7-20.

Winship, C., & Radbill, L. (1994). Sampling weights and regression analysis. *Sociological Methods & Research, 23*(2), 230-257.

Woldemicael, G. (1999). *Infant and child mortality in Eritrea: Levels, trends, and determinants.* Stockholm University,

Wolfinger, R. (1993). Covariance structure selection in general mixed models. *Communications in statistics-Simulation and computation, 22*(4), 1079-1106.

Wolfinger, R., & O'connell, M. (1993). Generalized linear mixed models a pseudo-likelihood approach. *Journal of statistical Computation Simulation, 48*(3-4), 233-243.

Wolfinger, R.D., & Lin, X. (1997). Two Taylor-series approximation methods for nonlinear mixed models. *Computational Statistics & Data Analysis, 25*(4), 465-490.

Wood, S.N. (2000). Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society: Series B (Methodological), 62*(2), 413-428.

Wood, S.N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian New Zealand Journal of Statistics, 48*(4), 445-464.

Wood, S.N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian New Zealand Journal of Statistics*

48(4), 445-464.

Wood, S.N. (2017). *Generalized additive models: an introduction with R*: CRC press.

Wood, S.N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association, 111*(516), 1548-1563.

Worldometers. (2019). http://www.worldometers.info/world-population/lesotho-population/. . Retrieved from http://www.worldometers.info/world-population/lesotho-population/. .20/07/2020

Wu, Z. (2005). Generalized linear models in family studies. *Journal of Marriage and Family, 67*(4), 1029-1047.

www.usaid.gov., https://2012-2017.usaid.gov/what-we-do/global-health/cross-cutting-areas/demographic-and-health-surveys-program. Retrieved from https://2012-2017.usaid.gov/what-we-do/global-health/cross-cutting-areas/demographic-and-health-surveys-program 20/09/2018.

Wyse, S.E. (2012). Advantages and disadvantages of surveys. *Retrieved February, 20*, 2014.

Xiang, D. (2001). *Fitting generalized additive models with the GAM procedure.* Paper presented at the SUGI Proceedings.

Xiang, D., & Wahba, G. (1996). A generalized approximate cross validation for smoothing splines with non-Gaussian data. *Statistica Sinica*, 675-692.

Yadav, S.K., Singh, S., & Gupta, R. (2019). Sampling Methods. In *Biomedical Statistics* (pp. 71-83): Springer.

Ying, G.-s., & Liu, C. (2006). *Statistical analysis of clustered data using SAS system.* Paper presented at the Proceedings of the north east SAS users group (NESUG) conference.

Yirga, A.A. (2018). *Statistical models to study the BMI of under five children in Ethopia.*

You, D., Hug, L., Ejdemyr, S., Idele, P., Hogan, D., Mathers, C., Gerland, P., New, J.R., & Alkema, L. (2015). Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. *The lancet, 386*(10010), 2275-2286.

Yu, Y., & Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association, 97*(460), 1042-1054.

Zou, K.H., O'Malley, A.J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation, 115*(5), 654-657.