

**UNIVERSITY OF GHANA, LEGON**



**MODELING MORTALITY AMONGST CHILDREN UNDER FIVE YEARS IN GHANA:**

**COMPARISM OF DIFFERENT MODELING TECHNIQUES**

**BY**

**SALIFU NANGA**

**(10443252)**

**THESIS SUBMITTED TO THE SCHOOL OF GRADUATE STUDIES, UNIVERSITY OF  
GHANA, LEGON, IN PARTIAL FULFILLMENT FOR THE AWARD OF THE MASTER  
OF PHILOSOPHY DEGREE IN STATISTICS**

**JUNE, 2015**

## DECLARATION

### Candidate's Declaration

This is to certify that this thesis is the result of research undertaken by Salifu Nanga towards the award of Master of Philosophy in Statistics in the Department of Statistics, University of Ghana.

Candidate's signature: ..... Date:.....

**SALIFU NANGA**

**(ID No. 10443252)**

### Supervisors' declaration

We hereby certify that this thesis was prepared from the candidate's own work and supervised in accordance with the guidelines on supervision of thesis laid down by the University of Ghana.

SIGNATURE: ..... DATE:.....

**DR. ANANI LOTSI**

**(Principal Supervisor)**

SIGNATURE: ..... DATE:.....

**DR. K DOKU – AMPONSAH**

**(Co – Supervisor)**

## ABSTRACT

Child mortality is regarded as one of the most revealing measures of society's ability to meet the needs of its people. The Millennium Development Goal 4 (MDG 4) advocates a reduction of under-five mortality rate by two-thirds between 1990 and 2015. The main objective of this study was to develop a validated set of statistical models and select the most appropriate model to predict mortality among children under five and to compare the influence of selected risk factors on the probability of death before the age of 5 years among children in Ghana. The study revealed that the  $k^{\text{th}}$  Nearest Neighbor was the most efficient in modeling Mortality in Children under five with a CCR of 83%. This is followed by Logistic Regression with a CCR of 81% and the least was Neural Network with a CCR of 80%. The highest educational level of mother, Age of mother at birth, Type of toilet facility used by family, alcohol consumption and the wealth index of family were discovered as the most important variables in predicting mortality amongst children under five in Ghana across all models.

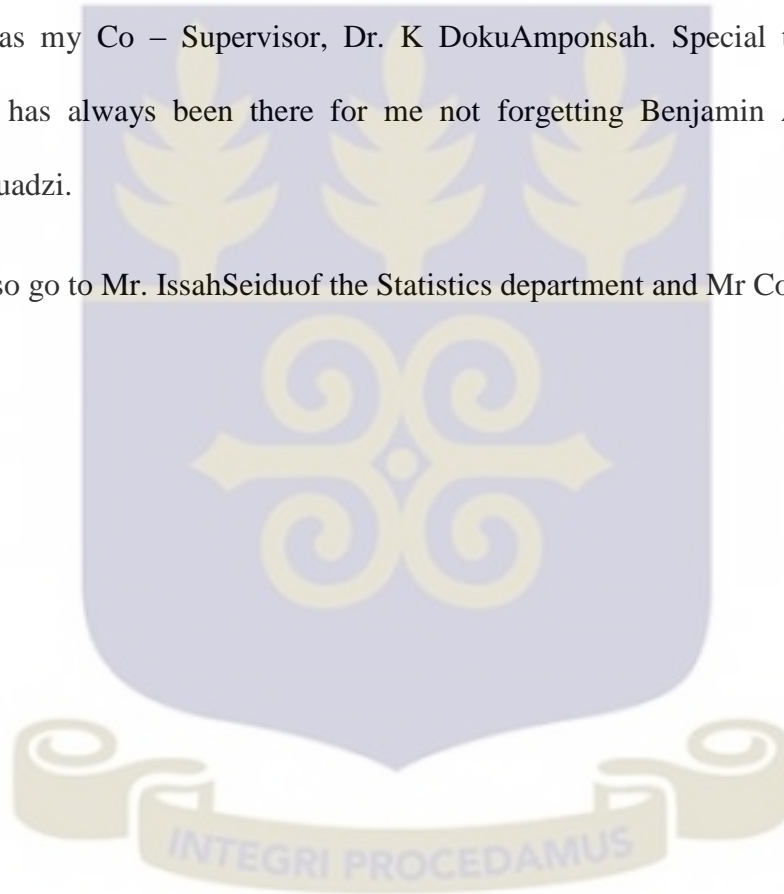
The study recommended that policy holders must ensure that every household has a place of convenience that is hygienic which has the tendency to prevent diseases like diarrhea which can result in the death of children under five. The government must also intensify public education on the dangers and effect of child mortality on society and also carry out measures to help reduce mortality significantly.

## ACKNOWLEDGEMENT

I convey my sincere thanks to all other lecturers in the Statistics department who provided me with valuable inputs regarding my work and helped me all along. I am also grateful to the non-teaching staff of the department for their support in diverse ways.

I am particularly grateful to my academic supervisor, Dr. Anani Lotsi for his supervision, time and support as well as my Co – Supervisor, Dr. K Doku Amponsah. Special thanks go to Lisa Adjoa Edzii who has always been there for me not forgetting Benjamin Ansa Acquaye and Benedictus Amekuadazi.

Special thanks also go to Mr. Issah Seidu of the Statistics department and Mr Collins Amartey.



## **DEDICATION**

I dedicate this work to my father, Mr. Seidu Nanga. May the Almighty Allah continue to shower his blessings on him.

This work is dedicated to all my family, friends and loved ones who helped me and always believed I could accomplish my goals. Thank you so much.



**Contents**

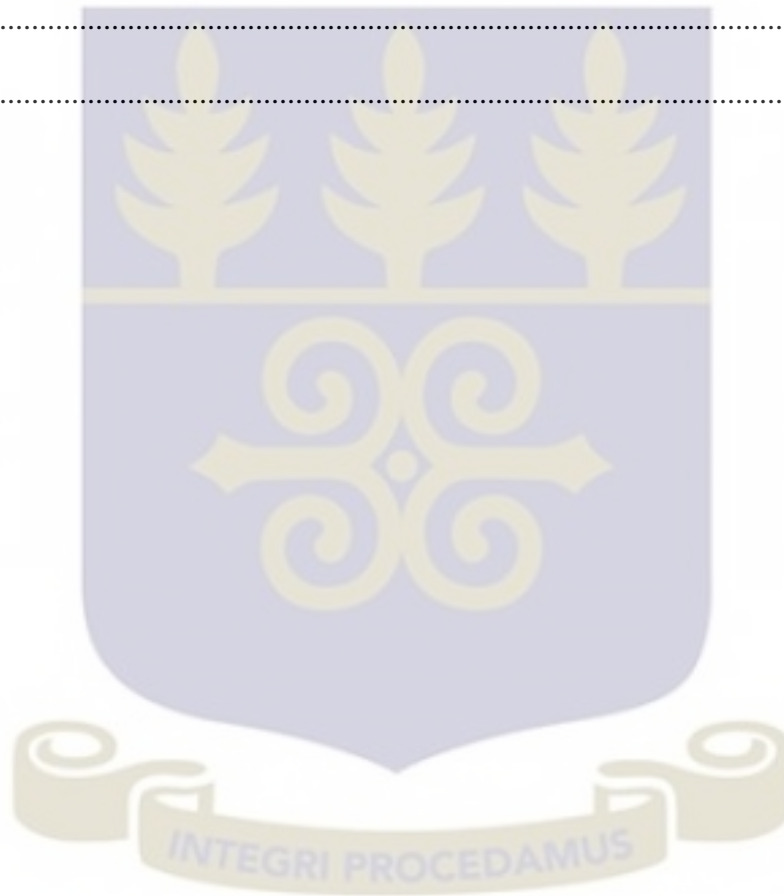
DECLARATION .....	i
ABSTRACT.....	ii
ACKNOWLEDGEMENT .....	iii
DEDICATION.....	iv
LIST OF TABLES .....	x
LIST OF FIGURES .....	xi
LIST OF ACRONYMS .....	xii
CHAPTER ONE.....	1
1.1 Background of Study.....	1
1.2 Problem Statement .....	6
1.3 Objective of the Study.....	6
1.4 Research Question.....	7
1.5 Scope of the Study.....	7
1.6 Significance of the Study .....	8
1.7 Organization of the Study .....	8
CHAPTER TWO .....	9
REVIEW OF RELATED LITERATURE .....	9
2.1 Introduction .....	9
2.2 Determinants of Mortality in Children.....	14
2.2.1 Marital Status and Child Mortality.....	16
2.2.2 Maternal Education and Child Mortality.....	17
2.2.3 Maternal Age and Child Mortality .....	20
2.2.4 Environmental Contamination and Child Mortality.....	20
2.2.5 Environmental Contamination and Child Mortality.....	21
2.2.6 Sex of the Child and Child Mortality .....	22

2.2.7 Place of Residence of the Mother and Child Mortality .....	24
2.2.8 Household Wealth Index and Infant Mortality .....	24
2.2.9 National Health Insurance .....	25
2.3 Statistical and Data Mining Techniques.....	26
CHAPTER THREE .....	36
METHODOLOGY .....	36
3.1 Introduction .....	36
3.2.0 Data Source.....	36
3.3 Logistic Regression .....	37
3.3.1.1 What Logistic Regression Predicts .....	38
3.3.1.2 Level of Measurement Requirements .....	38
3.3.1.3 Assumptions .....	38
3.3.1.4 Sample Size Requirements.....	39
3.3.2 Methods of Including Variables .....	39
3.3.3 Maximum Likelihood (ML) Estimation .....	39
3.3.4 Overall Test Of Relationship.....	43
<b>3.3.5 The Logistic Function</b> .....	43
3.3.6 The Logit Transformation .....	44
3.3.7 The Logistic Model .....	45
3.3.8 Hosmer Lemeshow (H-L) Goodness Of Fit Statistic .....	47
<b>3.11 Wald Statistic</b> .....	49
3.3.9 Odds Ratio .....	50
3.3.9.1 Definition of Odds Ratio .....	50
3.3.9.2 Derivation of the Odds Ratio Formulae .....	51
3.3.10 The Likelihood Ratio (LR) Test .....	53

3.3.11	Regression Coefficients .....	53
3.3.12	P-Value .....	54
3.3.13	Overall Model Fit .....	54
3.3.14	Study Assumptions .....	55
3.3.15	Omnibus Tests of Model Coefficients .....	55
3.3.17	Hosmer And Lemeshow Goodness Of Fit Test .....	56
3.3.18	Model Summary .....	56
3.3.19	Explanation Of Variables In The Model .....	56
3.4	Neural Network .....	57
3.4.1	Architecture .....	61
3.4.2	Hidden Layers.....	62
3.4.3	Output Layer and Output.....	62
3.4.4	Network Performance.....	62
3.4.5	Classification of results .....	63
3.4.6	Independent variable importance analysis.....	63
3.4.7	Probabilities and Pseudo-Probabilities .....	64
3.5	The K-Nearest-Neighbor Method .....	65
3.5.1	Classification .....	66
3.5.2	Regression .....	68
3.5.3	Technical Details .....	69
3.5.4	Cross-Validation.....	70
3.5.5	Distance Metric.....	71
3.5.6	Euclidean metric .....	71
3.5.7	City block metric .....	72
3.5.8	<i>k</i> -Nearest Neighbor Predictions.....	72

3.5.9 Distance Weighting .....	72
3.5.10 Neighbors.....	74
3.5.11 Number of Nearest Neighbors (k) .....	74
3.5.12 Feature Selection .....	74
3.5.13 Stopping Criterion .....	75
3.5.14 Minimum change in absolute error ratio .....	75
3.5.15 Partitions.....	75
3.5.16 Cross-Validation Folds.....	76
3.5.17 Output.....	76
3.5.18 Variable Importance .....	78
3.5.19 Classification Table .....	78
3.5.20 Error Summary .....	78
3.6 Comparison of Models .....	79
<b>CHAPTER FOUR.....</b>	<b>81</b>
<b>RESULTS AND DISCUSSION .....</b>	<b>81</b>
4.1 Introduction.....	81
4.2 Preliminary Analyses .....	81
4.3 Logistic Regression Model.....	85
4.4 Neural Network.....	90
4.5 Nearest Neighbor Analysis.....	96
4.6 Comparison of Models .....	102
<b>CHAPTER FIVE .....</b>	<b>105</b>
<b>SUMMARY, CONCLUSIONS AND RECOMMENDATIONS.....</b>	<b>105</b>
5.1 Introduction.....	105

5.2.0 Summary .....	105
5.2.1 Logistic Regression .....	105
5.2.2 Neural Networks.....	105
5.2.3 K Nearest Neighbor .....	106
5.3 Conclusion.....	107
5.4 Recommendations .....	109
REFERENCES .....	110
APPENDIX.....	120

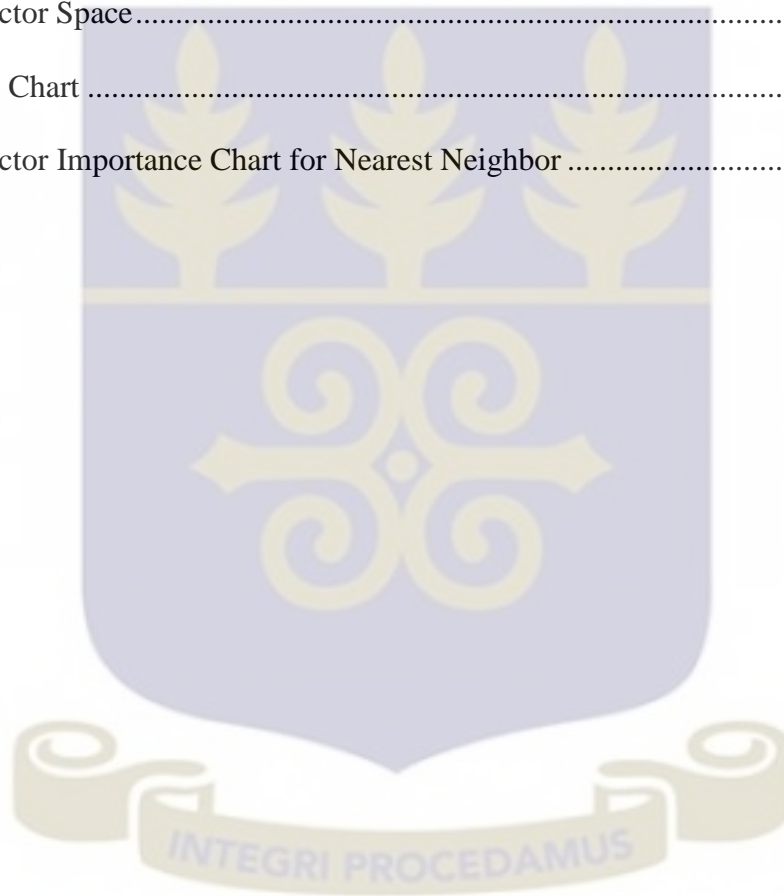


**LIST OF TABLES**

<b>Table 4. 1: Frequency Distribution of Variables used in the study .....</b>	<b>82</b>
Table 4. 2: Pearson Chi-Square Tests .....	83
Table 4. 3: Age of mother at birth.....	84
Table 4. 4: Case Processing Summary.....	85
Table 4. 5: Model Summary .....	85
Table 4. 6: Classification Table .....	86
Table 4. 7: Omnibus Tests of Model Coefficients .....	86
Table 4. 8: Variables in the Equation of the Logistic Regression Model .....	87
Table 4. 9: Hosmer and Lemeshow Test .....	89
Table 4. 10: Case Processing Summary.....	90
Table 4. 11: Model Summary .....	92
Table 4. 12: Classification Table .....	92
Table 4. 13: Area under the Curve.....	94
Table 4. 14: Independent Variable Importance.....	94
Table 4. 15: Case Processing Summary.....	96
Table 4. 16: Error Summary for Nearest Neighbor .....	97
Table 4. 17: 4 Nearest Neighbors and Distances .....	99
Table 4. 18: Summary of the Peers Chart.....	99
Table 4. 19: Classification Table for Nearest Neighbor .....	101
Table 4. 20: Comparison Table.....	102
Table 4. 21: Correct Classification rate Table .....	103

## LIST OF FIGURES

Figure 4. 1:Feedforward architecture with one hidden layer .....	91
Figure 4. 2: ROC Curve of the Neural Network Model .....	93
Figure 4. 3: Normalized Importance Chart .....	95
Figure 4. 4: K Selection Error Log Chart .....	96
Figure 4. 5: Predictor Space .....	97
Figure 4. 6: Peers Chart .....	98
Figure 4. 7: Predictor Importance Chart for Nearest Neighbor .....	100



## LIST OF ACRONYMS

CART	Classification And Regression Trees
CCR	Correct Classification Rate
GDHS	Ghana Demographic and Health Survey
H-L	Hosmer and Lemeshow's
k-NN	k-Nearest Neighbor
LDA	Linear Discriminant Analysis
LRLogistic Regression	
LRT	Likelihood Ratio Test
MDG 4	Millennium Development Goal 4
ML	Maximum Likelihood
MLP	Multilayer Perceptron
NGOs	Non-Governmental Organizations
NN	Neural Network
OR	Odds Ratio
PDA	Predictive Discriminant Analysis
QDA	Quadratic Discriminant Analysis
ROC	Receiver Operating Characteristic
U5MR	Under Five Mortality Rate
UN	United Nations
WHO	World Health Organization

## CHAPTER ONE

### 1.1 Background of Study

Under five mortality rate (U5MR) is defined as the likelihood that a newborn will die before reaching the age of five years. Nearly 10 million children worldwide die before their fifth birthday every year, with almost all of such deaths occurring in developing countries; sub-Saharan Africa accounts for almost 50% of these deaths (Black et al., 2003; Claeson et al., 2003, WHO, 2007; UNICEF, 2008).

Child mortality is regarded as one of the most revealing measures of how well a society is meeting the needs of its people (Buckley, 2003). Millennium Development Goal 4 (MDG 4) calls for reducing the under-five mortality rate by two-thirds between 1990 and 2015 (Unicef, 2014).

The worldwide toll of under-five deaths over the past two decades is quite overwhelming: between 1990 and 2013, 223 million children died before age five. Globally, the 49 percent decline in the under-five mortality rate since 1990 is still far below the two-thirds reduction required to reach the Millennium Development Goal 4 target. If current trends continue, only three regions—Eastern Asia, Latin America and the Caribbean, and Northern Africa—will achieve MDG 4 by 2015. The rate of decline in under-five mortality in all other regions remains insufficient to achieve MDG 4.

A child's risk of dying before age five increases if she or he is born in a remote rural area, into a poor household or to a mother with no education (UNICEF, 2010).

Previous research has indicated that various factors influence a child's health and survival, including place of residence, breastfeeding, place of delivery, access to postnatal care, and

maternal age and education (Doctor HV,2011). Geographic differences in maternal literacy levels and sociocultural practices are linked to the variations in child mortality rates within countries (Black et al,2003) . The likelihood of under-five mortality has also been linked to place of delivery, with evidence indicating that women who deliver at health facilities have a lower probability of reporting child death compared with those delivering in home settings (Doctor HV,2011).

Access to postnatal care has also been associated with a drop in under-five mortality, with a study in Bangladesh showing that postnatal home visits within the first 2 days after birth by skilled healthcare workers was significantly associated with a lower likelihood of child death (Baqui,2009).Studies conducted in the developing countries have found that maternal education and age are important determinants of child mortality (Deribew et al., 2007). Evidence shows that child mortality rates are higher among less educated mothers compared with mothers who have higher levels of education (UN, 2010).

The importance of education, particularly mother's education, has been confirmed in many subsequent studies (Murthi et al., 1995; Dre`ze and Murthi, 2001).To the extent that, education improves an individual's ability to undertake these changes, more educated mothers will have healthier babies (Meara, 2001; Currie and Moretti,2003). Mother's employment status is also considered an important factor affecting neonatal, infant and child mortality (Arriaga and Hobbs, 1982). The mother's work status determines the amount of time and care a mother can give to her child, and it may determine the amount of resources (income) available to the mother and thus her access to various goods and services.

Ghana was ranked forty-seventh (47th) in the world in terms of under five mortality rate in 2002, with a rate of 100 per 1000 live births (down from 126 per 1000 live births in 1990) and an

Infant Mortality Rate (IMR) of 57 per 1000 live births (UNICEF, 2004). However, the rates increased slightly to 111 per 1000 live births and 64 per 1000 live births for under-five and infant mortality respectively for the year 2003 and in 2008, decreased to 80 and 50 per 1000 live births respectively (GSS/GHS//ORC Macro, 2008 preliminary results).

To achieve the 4<sup>th</sup> Millennium Development Goal (MDG), Ghana has to reduce her U5MR rate to 40 per 1000 live births by 2015 (Bryce et al., 2008). The annual rate of U5MR reduction in Ghana (1990-2006) is 0.09%. Between 2007 and 2015, Ghana has to achieve an annual rate of reduction of 12.2% in order to achieve the MDG 4 (Countdown Coverage Writing Group, 2008)

The purpose of this study is to assess what factors are most important in determining mortality in children under five years in Ghana comparing three different modeling techniques, namely , Logistic Regression and two data mining techniques namely, non-parametric discriminant analysis (Kth Nearest Neighbor) and Neural Networks.

Logistic regression is statistical technique used to model data in which the target or dependent variable is binary. The main objective is to develop a regression type model relating the binary variable to the independent variables. It can also be used to examine the variation in the dependent variable that can be explained by the independent variables, to rank the independent variables based on their relative importance in predicting the target variable, and to determine the interaction effects among independent variables. Rather than predicting the values of the dependent variable, logistic regression estimates the probability that a dependent variable will have a given value.

Data mining is the process of selecting, exploring, and modeling large amounts of data to uncover new trends and patterns in massive databases. These analyses lead to proactive decision making and knowledge discovery in large databases by stressing data exploration to thoroughly studying the structure of data and to checking the validity of statistical models that fit. The two data mining techniques considered in this study are Neural Networks and Non parametric discriminant analysis (kth Nearest Neighbor).

The term neural network applies to a loosely related family of models, characterized by a large parameter space and flexible structure, descending from studies of brain functioning. As the family grew, most of the new models were designed for non-biological applications, though much of the associated terminology reflects its origin. Specific definitions of neural networks are as varied as the fields in which they are used. While no single definition properly covers the entire family of models, for now, consider the following description (Haykin, 1998): A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

- Knowledge is acquired by the network through a learning process.
- Interneuron connection strengths known as synaptic weights are used to store the knowledge

In order to differentiate neural networks from traditional statistical methods using this definition, what is not said is just as significant as the actual text of the definition. For example, the traditional linear regression model can acquire knowledge through the least-squares method and store that knowledge in the regression coefficients. In this sense, it is a neural network. It can be argued that linear regression is a special case of certain neural networks. However, linear

regression has a rigid model structure and set of assumptions that are imposed before learning from the data. By contrast, the definition above makes minimal demands on model structure and assumptions. Thus, a neural network can approximate a wide range of statistical models without requiring that we hypothesize in advance certain relationships between the dependent and independent variables. Instead, the form of the relationships is determined during the learning process. If a linear relationship between the dependent and independent variables is appropriate, the results of the neural network should closely approximate those of the linear regression model. If a nonlinear relationship is more appropriate, the neural network will automatically approximate the “correct” model structure.

Discriminant analysis is one of the data mining techniques used to discriminate a single classification variable using multiple attributes. Discriminant analysis also assigns observations to one of the pre-defined groups based on the knowledge of the multi-attributes. When the distribution within each group is multivariate normal, a parametric method can be used to develop a discriminant function using a generalized squared distance measure. The classification criterion is derived based on either the individual within-group covariance matrices or the pooled covariance matrix that also takes into account the prior probabilities of the classes.

Non-parametric discriminant methods are based on non-parametric group-specific probability densities. Either a kernel or the k-nearest-neighbor method can be used to generate a non-parametric density estimate in each group and to produce a classification criterion. The performance of a discriminant criterion could be evaluated by estimating probabilities of misclassification of new observations in the validation data.

In this study, the results obtained by using logistic regression were compared with the results obtained by using neural networks and non - parametric discriminant analysis (Kth Nearest Neighbor).

### **1.2 Problem Statement**

Every year more than ten million children worldwide die before their fifth birthday, most of them during the first year. Under-five mortality rate varies globally from 0.3 percent in high developed countries like Sweden to an average of 17 percent in sub-Saharan Africa. In 2013, the World Health Organization's (WHO) report on child mortality estimated mortality in under five years for developed countries as 6 deaths per 1000 live births and 53 deaths per 1000 live births developing countries, as at the year 2012. This means developed countries are on course to achieving their MDG target of 5 deaths per 1000 live births. However, Sub-Saharan Africa has a long way to go to catch up with their developed counterparts. The subject of mortality in under five children has increasingly gained importance to the process of development of any country and Ghana is no exception.

There are varying factors that are considered to contribute to mortality in children under five. However the problem has been the magnitude of significance this factors play. Hence, it is important to develop an effective model to predict mortality in children under five in Ghana, taking into consideration the various factors.

### **1.3 Objective of the Study**

The objective of this study is, therefore, to compare the influence of selected risk factors on the probability of death before the age of 5 years among children in Ghana and also select the most appropriate model in predicting mortality among children under five.

Specifically, the study seeks to:

- (i) Determine what kinds of risk factors that are most important in predicting mortality in children under five years in Ghana.
- (ii) To predict the likelihood that a child under five will die or live, taking into consideration a set of given covariates or factors.
- (iii) Compare the results obtained with both statistical technique (Logistic Regression) and data mining techniques ( $K^{\text{th}}$ Nearest Neighbour and Neural Networks) to determine which of them is more adequate in terms of fit and predictive capacity.

#### **1.4 Research Question**

- (i) What kinds of risk factors that are most important in predicting mortality in children under five years in Ghana?
- (ii) What is the likelihood that a child under five will die or live, taking into consideration a set of given covariates or factors?
- (iii) How can a comparison of modeling techniques such as Logistic Regression, Neural Networks and  $K^{\text{th}}$  Nearest Neighbor be used in predicting mortality in children under five years in Ghana?

#### **1.5 Scope of the Study**

This study focuses on the 2008 Ghana Demographic and Health Survey (GDHS) and examines mortality among under-five children based on some socio-demographic indicators.

## **1.6 Significance of the Study**

Prediction of mortality in children has long been regarded as an important research topic in many academic disciplines for a number of reasons. First, predictive models can enable demographers, and other stake holders (policy makers, planners, academicians and NGOs) to predict correctly expected number of deaths and then take some proactive measures.

This study will also reveal some of the determinants of mortality among under five children in Ghana. Mortality in under five children is a major concern particularly in developing nations, where the effect is greatest. The World Health Organization (WHO) report on child mortality (2013) estimated mortality in children under five years at 3.5 million deaths a year.

This study will also enable us ascertain which factors play significant roles in determining mortality in children under five years in Ghana. It will also enable us to select the most appropriate technique in modeling mortality in children under five.

Finally, recommendations based on the findings would serve as tools for further research into the subject matter, to ensure that acceptable and lasting solutions are found. The results of the study will also be a yardstick upon which corrective measures will be taken to help control mortality in children under five and also to sensitize the general public.

## **1.7 Organization of the Study**

The study is organized into five (5) chapters. The first chapter consists of introduction and gives a background to the study. The chapter also discusses the statement of the problem, objectives and the research questions, significance and scope of the study. The second chapter focuses on literature on mortality in children under five years. Chapter three details the methodology

adopted for the study. It indicates the, sample, research design and instrument for data collection as well as the statistical techniques used to assist in achieving the research objectives. The fourth chapter covers data analysis and presentation of the study. The fifth and final chapter constitutes the conclusions and recommendations based on the findings.



## **CHAPTER TWO**

### **REVIEW OF RELATED LITERATURE**

#### **2.1 Introduction**

Data mining applies advanced data analysis techniques to explore the data and identify useful and unexpected patterns and rules that provide relevant knowledge for predicting future

outcomes. It can be defined as a process of exploring and analysing large amounts of data with a specific target on discovering significantly important patterns and rules. Data mining helps finding knowledge from raw, unprocessed data. Using data mining techniques allows extracting knowledge from the data mart, data warehouse and, in particular cases, even from operational databases ( Lungu and Bâra,2012).

Data mining became a vital discovery process for various domains including business where data mining can be used in retail, banking or insurances, for purposes like customer segmentation, market basket analysis or fraud detection. It can also be applied in medicine, where it can help with diagnosis prognosis or treatment planning, bioinformatics, genetics, electrical power or astronomy. It is also applied in education where data mining can be used for predicting student performance, understanding student behavior or student segmentation.

It has many successful applications, such as business intelligence, Web search, bioinformatics, health informatics, finance, digital libraries, and digital governments (Han et al,2011).

Data mining methods are incorporating techniques from statistics, artificial intelligence, machine learning and database systems. According to the Gartner Group (The world's leading information technology research and advisory company), data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques(Larose, 2005).

Therefore, the main purpose of data mining consists in applying advanced data analysis techniques on large volumes of data stored in databases, data marts or data warehouses, in order to discover significant patterns, trends, rules, relations and correlations contained within the data,

information, which would be impossible or very difficult to observe using other data analysis techniques.

Neural network is one of the data mining techniques used in exploring data. True neural networks are biological systems that are used to detect patterns and make predictions. The artificial ones are computer programs implementing sophisticated pattern detection and machine learning algorithms on a computer to build predictive models from large historical databases. It derives its name from a historical development which started off with the premise that machines could be made to “think” if scientists found ways to mimic the structure and functioning of the human brain on the computer. Thus historically neural networks grew out of the community of artificial Intelligence rather than from the discipline of statistics. Despite the fact that scientists are still far from understanding the human brain comprehensively let alone mimicking it, neural networks that run on computers can perform some of the tasks that people can do.

It is difficult to say precisely when the first “neural network” on a computer was built. During World War II, a groundbreaking paper was published by McCulloch and Pitts(2003), which first outlined the idea that simple processing units (like the individual neurons in the human brain) could be connected together in large networks to create a system that could solve difficult problems and display behavior that was much more complex than the simple pieces that made it up. Since that time much advancement has been made in discovering ways to apply artificial neural networks to real world prediction problems and in improving the performance of the algorithm in general. In the early 1980's, researchers showed renewed interest in neural networks. Recent work includes Boltzmann machines, Hopfield nets, competitive learning models, multilayer networks, and adaptive resonance theory models. In many respects the greatest breakthroughs in neural networks in recent years have been in their application to more

mundane real world problems like customer response prediction or fraud detection rather than the loftier goals that were originally set out for the techniques such as overall human learning and computer speech and image understanding.

Nearest Neighbor prediction techniques are among one of the oldest techniques used in data mining. It is a prediction technique that is quite similar to clustering. Its essence is that in order to predict what a prediction value is in one record, it looks for records with similar predictor values in the historical database and uses the prediction value from the record that is “nearest” to the unclassified record.

Statistical methods on the other hand, can be described as being characterized by the ability to only handle data sets, which are small and clean, which permit straightforward answers via intensive analysis of single data sets, which are static, which were sampled in an iid (variables are independent and identically distributed if each has the same probability distribution as the others and all are mutually independent) manner, which were often collected to answer the particular problem being addressed and often which are solely numeric. Literature shows that a variety of statistical methods have been used in the past for classification tasks. Logistic regression is one of the most popular statistical methods used in in classifying data. It is important to understand that the goal of an analysis using logistic regression is the same as that of any model-building technique used in statistics: to find the best fit and most parsimonious. What distinguishes a logistic regression model from a linear regression model is that of the outcome variable. In the logistic regression model, the outcome variable is binary or dichotomous. The difference between logistic and linear regression is reflected both in the choice of a parametric model and in the assumptions. Once this difference is accounted for, the methods employed in an analysis using logistic regression follow the same general principles used in

linear regression analysis (Cizerk&Fritzgerald, 1999). The under-five mortality rate is an important indicator of child well-being, including health and nutrition status. It is also a key pointer of the coverage of child survival interventions and, more broadly, of social and economic development.

Millennium Development Goal 4 (MDG 4) calls for reducing the under-five mortality rate by two-thirds between 1990 and 2015. Substantial progress has been made in reducing the rate 49 percent, from 90 (89, 92) deaths per 1,000 live births in 1990 to 46 (44, 48) in 2013. Since 1990 almost 100 million children under age five roughly the current population of the Philippines have been saved.

The world is also reducing under-five mortality faster than at any other time during the past two decades. The global annual rate of reduction has steadily accelerated since 1990–1995 more than tripling from 1.2 percent to 4.0 percent in 2005–2013. Despite these gains, child survival remains an urgent concern. The toll of under-five deaths over the past two decades is staggering: between 1990 and 2013, 223 million children worldwide died before their fifth birthday more than today's population of Brazil, the world's fifth most populous country. Progress has been insufficient, and the MDG 4 target risks being missed at the global level. To achieve MDG 4 on time, the global annual rate of reduction in under-five mortality would have to rise to 20.8 percent for 2013–2015, much higher than the 4.0 percent achieved over 2005–2013. At the country level, historical trends show that progress for most countries has been too slow and that only 12 of the 60 countries with high under-five mortality rates (at least 40 deaths per 1,000 live births) are on course to achieve MDG 4 if current trends continue. It is unacceptable that every day 17,000 children still die before their fifth birthday, mostly from preventable causes and

treatable diseases, even though the knowledge and technologies for lifesaving interventions are available (WHO,2014).

## **2.2 Determinants of Mortality in Children**

Under-five mortality decline has improved from 1.2% per year between 1990 and 1995, to 3.9% per year between 2005 and 2012 (UNICEF, WHO, World Bank, 2013). In spite of this substantial drop in global child mortality rate, about 6.6 million children still die every year before their fifth birthday worldwide which implies 18,000 under-five children die each day (UNICEF, WHO, World Bank, 2013). There are huge disparities in child mortality among low and middle income countries and the industrial world with Sub-Saharan Africa and South East Asia carrying the highest burden of under-five mortality (Rahman et al. , 2010).

Previous studies have revealed that childhood mortality varies due to the variations of associated characteristics of the parents as well as children under five (Preston,1995). Focusing on 28 developing countries mostly in Asia and Latin America, Hobcraft et al. (1994) revealed that mother's and husband's education, their work status and their residence were more or less associated with child survival. Da Vanzo et al.,(1993) in Malaysia revealed a higher risk of death to children born to mothers below 18 and above 40 years of age. Children in sub-Saharan Africa have the highest risk of death in the first month of life and are still leading in under-five mortality rates with one in every nine children dying before their fifth birthday as of 2011 (You et al.,2014). It is worth noting that the year 2008 recorded a rate of one in seven children (144 per 1000 live births) dying before their fifth anniversary with the highest levels occurring in West and Central Africa. Within 34 countries where under-five mortality exceeded 100 out of 1000 live births in 2008, all except one are in sub-Saharan Africa (You et al., 2010). The rate of

improvement in child survival in Sub-Saharan Africa is unsatisfactory to meeting United Nations Millennium Development Goal 4 of reducing under-five mortality rate by two-thirds between 1990 and 2015 as the region has the highest risk of death in the first month of life and is among the regions showing the least progress globally (UNICEF, WHO, World Bank, 2013). In order to hasten the decline in under-five mortality rate, specific proven interventions would have to target important causes of child death (Jones et al., 2003). Since no single factor can account for the high child mortality, in developing these interventions there is the need to understand the multiplicity of factors that determine child mortality especially in resource underprivileged settings (Adhikari et al., 2010).

Previous studies have shown that various factors influence child health and survival including place of residence, mothers' age, mother's education, place of delivery, birth order, sex of child, religion of parents, household headship and household socio economic status (Adhikari et al., 2010; Ettarh et al., 2012). Though poverty is well acclaimed as an essential factor influencing child mortality (Swenson et al., 1993; Awoonor et al., 2013), findings on the effect of household socio economic differentials on child mortality have been mixed. A study in parts of rural Ghana and another in Tanzania did not find any significant effect of household socio economic status on child mortality (Welaga et al., 2013). A study using Nigeria Demographic and Health Survey for 2008, revealed that relatively prosperous households were less likely to experience child death than the poorest households in rural Nigeria (Doctor, 2013).

Existing literature has documented mixed results, such as place of delivery, birth order and sex of child. For instance, abundance of evidence suggest that women who deliver at health facilities have a lower chance of child death as compared to those who deliver at home due to the use of skilled delivery at health facilities and the none existence of such at home (Ettarh et al., 2012).

However, studies in South Africa and parts of Nigeria suggest that place of delivery does not have significant effect on either perinatal or under-five mortality (Worku,2009).

In Ghana, just as in many other countries in Sub-Saharan Africa with low socio economic development, under-five mortality is relatively high with a recent reported national figure of 90 per 1000 live births (Ghana Statistical Service,2013). Moreover within Ghana, disparities exist in under-five mortality rates between regions. In the more resource rich regions of southern Ghana, under-five mortality rate ranges from 75 per 1000 live births while in the most impoverished and deprived regions of the north such as the Upper East Region it is as high as 128 per 1000 live births (Ghana Statistical Service,2013). There is the need for concerted efforts if we hope to achieve the desired improvements in under-five mortality. In developing countries maternal, demographic, and socioeconomic factors have been found to be important determinants of childhood mortality (Mosley and Chen, 1984; Mutunga, 2007; Rutstein, 2000; Gyimah and Fernando, 2002). In addition, rapid urban growth often has outpaced the provision of safe water and sanitation, with crowded living conditions facilitating the spread of diseases that can affect child survival (Rutstein, 2000, Mishra and Retherford, 2007). In India, a study conducted by Mishra in 2007 revealed that the socioeconomic status of mothers is salient in the reduction of childhood morbidity and mortality from health hazards within the household.

### **2.2.1 Marital Status and Child Mortality**

In a study carried out in 11 countries in Sub-Saharan Africa in 2013, survival analysis techniques were used to estimate the likelihood of becoming a single mother over women's life course and investigate the relationship between single motherhood and child mortality. In nine countries, being a formerly married (divorced or not re – marry after spouse's death) mother was associated with a significantly higher risk of dying (odds ratios range from 1.29 in Zambia to 1.75 in

Kenya) relative to having married parents. Children of divorced women typically had the poorest outcomes. These results highlight the vulnerability of children with single mothers and suggest that policies aimed at supporting single mothers could help to further reduce child mortality in sub-Saharan Africa (Clark et al., 2013).

A study carried out in Sweden found that growing up in a single-parent family has disadvantages to the health of the child. Lack of household resources plays an important part in increased risks. However, even when a wide range of demographic and socioeconomic circumstances are included in multivariate models, children of single parents still have increased risks of mortality, severe morbidity, and injury. (Weitof et al., 2003).

### **2.2.2 Maternal Education and Child Mortality**

Overwhelming micro- and macro-level evidence suggests a negative association between maternal education and child mortality (see Cochrane, Leslie & O'Hara 1980; Cleland & van Ginneken, 1988; Cleland, 1990). The single biggest factor, by far, in reducing the rate of death among children younger than five is greater education for women. In all countries worldwide, whether females increase schooling from 10 years to 11, for instance, or two years to three, child mortality declines, according to a recent study by the Institute for Health Metrics and Evaluation at the University of Washington in 2011. Women with more education tend to have smaller families, in part because of increased employment opportunities and better knowledge about contraception; fewer children in a family increases the chances that an infant will survive. More education also helps women make better decisions about many health and disease factors such as prenatal care, basic hygiene, nutrition and immunization.

A mother's education level has a huge, if not direct, effect on the health of her children. That relationship, observed in many small studies in rich nations, turns out to be true everywhere on the globe, according to a new study conducted by Unicef. Half the reduction in child mortality over the past 40 years can be attributed to the better education of women, according to the analysis published in the journal *Lancet* in 2007. For every one year, increase in the average education of reproductive age women, a country experienced a 9.5 percent decrease in the child deaths. It may be useful to recapitulate the key routes of causation identified by the earlier studies before attempting to review the possible influence of female labour force participation on child mortality. The surveys indicate that the way in which maternal education influences child mortality is fairly complex and has three different facets. First, the observed relationship between maternal education and child mortality could partly be due to certain independent factors associated with education, such as different fertility behaviour and higher economic and social status, which reduce mortality risk. Secondly, education itself can have an independent influence on child mortality by promoting better child care practices at home and more rigorous use of preventive and curative healthcare. Finally, certain extraneous factors may either enhance or suppress the overall strength of the association between child health and maternal education. What is observed is an amalgam of these three routes of causation. Part of the observed association between maternal education and child mortality may be due to certain independent factors connected with educated mothers. Different fertility behaviour and better socioeconomic status are often cited as important associated factors. The fertility behaviour of educated mothers minimizes the child mortality risk associated with birth as they tend to have children when they are neither too young nor too old; they also may be better at spacing their births. All these factors are known to reduce the child mortality risk. Similarly, with higher education, women are likely

to have higher incomes and better social status either through their direct participation in the labour market or through the higher probability of being married to wealthier men. Studies using World Fertility Survey data clearly show that education has a strong favourable effect on child mortality independent of the influence of different fertility behaviour and better socioeconomic status (Hobcraft, McDonald and Rutstein 1984; Cleland and van Ginneken 1988).

Mother's education affects child survival in two main ways: through better child-care practices and higher standards of hygiene at home, and more rational and greater use of preventive and curative medical services (Mosley and Chen 1984; Cleland and van Ginneken 1988). The first way is known as the 'household production of health' in the health economics literature (Grossman 1972).

It is hypothesized that the effectiveness with which basic child-health-promoting inputs, such as personal hygiene, prenatal and postnatal care, and feeding practices are combined, improves with the level of education of the mother. It is also argued that education gives greater independence to the mother which will help her take child-health-promoting decisions without any hindrance (Caldwell 1986).

There is considerable empirical evidence, from the less developed countries in all parts of the world, that the propensity to use preventive and curative health services for self and children is high among educated mothers. Educated mothers are also found to have superior knowledge of diseases and they seek timely treatment more often (see Cleland & van Ginneken, 1988). However, some studies deny superior health-care knowledge on the part of educated mothers, particularly among those with lower levels of education (Caldwell, Reddy & Caldwell, 1983; Lindenbaum, Chakraborty & Elias, 1985). To sum up, the available evidence indicates a strong

and independent association between mother's education and child health, but the exact mechanisms through which it operates are not yet clear.

### **2.2.3 Maternal Age and Child Mortality**

Several studies revealed that, age of the mother is statistically significant in explaining infant and child mortality. For example; mothers who had their first birth before age 20 are about 2.4 times more likely to have child death compared to those who had theirs at ages 20-34 (Ntimba and Mbago, 2005). It is generally expected that children born to young mothers (aged less than 20 years) and those born to older mothers (aged 40-49 years) should have higher mortality than those born to mothers aged 20-39 years (Mustafa & Odimegwu, 2003, Kembo & Ginneken, 2009).

### **2.2.4 Environmental Contamination and Child Mortality**

Improved sanitation lowers mortality by the mechanism of less exposure of the children to a contaminated environment. In turn, this makes them less vulnerable to disease and eventually death. Various studies have established the effects of type of toilet facility on infant mortality. Incidence of infectious diseases such as diarrhoea is seen to be influenced by the state of the environment. This type of toilet facility may facilitate pollution and contamination of the environment and consequently affect child mortality. According to a study conducted by Wak, (2002) in the Kassena-Nankana district in Ghana, it was found that, infant mortality is experienced by children whose compounds have no toilet facilities while those who use water closet or pan toilet latrine experience the lowest infant deaths. This study further indicated that children born to mothers whose drinking water is from unprotected sources, experience high child deaths than children born by mothers whose drinking water is from protected sources.

According to UNICEF, poor hygiene, lack of access to safe drinking water, and sanitation causing cholera and diarrheal diseases are responsible for the death of 1.5 billion children each year (UNICEF, 2008). Despite the fact that water and sanitation are seen as central to policies to reduce child mortality (a Lancet editorial in 2007 described adequate sanitation as ‘the most effective public health intervention the international community has at its disposal’), there have been few comparable studies attempting a more comprehensive analysis of the health benefits of improving drinking water and sanitation in developing countries (Günther, 2007). Günther, Günther and Hill (2007) conducted one of the most comprehensive analyses on child health, water, and sanitation. The authors merged all the DHS datasets available for seventy countries over the period 1986 to 2007. Even though the estimated effect on improved water and sanitation is smaller than estimations done by other studies, they still found a positive effect in the reduction of mortality, as well as a lower risk of diarrhea, and stunting. However, the authors also found that the positive results of clean water are more subtle and affect only children between 1 and 12 months. Across countries, access to safe water generally had a more important impact on child health in rural areas, while access to improved sanitation improved child survival in urban areas (Fuentes, Pfuetze & Seck, 2006). Trussell and Hammerslough (1993) found that improved latrines reduced child mortality in Sri Lanka, but that the source of water supply was insignificant. In Malaysia, Ridder and Tunali (2011) did not find any impact of access to piped water and toilet facilities on child mortality.

### **2.2.5 Environmental Contamination and Child Mortality**

Improved sanitation lowers mortality by the mechanism of less exposure of the children to contaminated environment. In turn, this makes them less susceptible to disease and eventually death. Various studies have established the effects of type of toilet facility on infant mortality.

Incidence of infectious diseases such as diarrhoea is seen to be influenced by the state of the environment. This type of toilet facility may facilitate pollution and contamination of the environment and consequently affect infant mortality. According to unpublished study conducted by Wak, (2002) in the Kassena-Nankana district in Ghana, it was found that, child mortality is experienced by children whose compounds have no toilet facilities while those who use water closet or pan toilet latrine experience the lowest infant deaths. This study further indicated that children born to mothers whose drinking water is from unprotected sources, experience high child deaths than children born by mothers whose drinking water is from protected sources.

#### **2.2.6 Sex of the Child and Child Mortality**

There has also been studies to determine the effect of gender differentials on mortality in children under five years. A report covering data over a decade was produced by the United Nations, in over a decade. The results of this study highlighted the importance of estimating childhood mortality rates separately by sex. In many areas of the world, advances in survival appear to be accruing relatively equitably to girls and boys. In many of the less developed regions, girls' past disadvantage in mortality at ages 1-4 appears to be easing. However, this is not universally the case. In China and India there is evidence that girls are not benefiting as much as boys from the national trends of mortality decline. On the other hand, several countries had findings suggesting a greater than expected degree of male disadvantage in survival during infancy (UN,2010).In normal conditions, i.e. in the absence of gender discrimination, female child mortality is lower than their male counterpart mainly because the female child is biologically stronger than the male child. But over the course of time from infancy to childhood,

the female child is discriminated in terms of nutritious food, proper health care etc. and that results in a higher female mortality during childhood.

Existing demographic literature is full of studies that have dealt with postnatal discrimination against female children and its impact on excess female child mortality (Miller, 1992; Gupta, 1990; Sen, 1988; Kishor, 2004). Sex differential in early childhood mortality or in other words, female disadvantage in child survival is well documented in most of the South Asian countries. In India, this has been a subject of great concern, especially in recent times. This female disadvantage remains a great challenge for achieving gender equity. Eliminating such differences may substantially reduce overall child mortality (Arokiasamy, 2004).

Child mortality indicators over the last three decades reveal that although levels of child mortality have declined, sex differentials in mortality particularly during the early childhood have actually widened in many countries of South-central Asia. In this region, which includes countries like China and India, excess female mortality in childhood is estimated at about 250,000 preventable deaths among girls under age 5 (United Nations, 2010).

Bhat (2002) in his paper cited some recent articles, which argued that the falling female-male ratio can be explained as a lagged effect of sex differences in mortality at young ages persisting over a long period of time. A child's sex has been shown to affect the probability of infant and child mortality: owing to biological factors, male infants have a higher risk of mortality during the first year of life, as highlighted for example in the report by WHO, (2007). In addition, differential treatment of boys and girls, owing to cultural and socioeconomic factors, may also be expected to affect the chances of survival during childhood (Kaldewei, 2010). Male children generally experience slightly higher mortality during infancy and childhood than the female

children. Evidence from demographic literature attributes the excess mortality among male children mostly to their higher biological risks during the first year of life. For example, the study conducted by Goro, (2007) in the three region of Ghana; namely Northern region, Upper West and Upper East. The study found that, infant mortality for males in the Upper East region was higher (84 per 1,000 live births) than females (74 per 1,000 live births).

### **2.2.7 Place of Residence of the Mother and Child Mortality**

The variation in climate and vegetation can explain differences in morbidity and mortality by area, place or zone of residence. For example a child given birth to by a mother residing in a very dry land where crop cultivation is not easy, may have a problem in obtaining body protective foods especially if the family is not rich enough to buy from the shops. According to the study conducted by Sello, (2003), using Lesotho DHS (2001), showed that children born in mountainous areas were more likely to die compared to children born in lowland areas. The author argued that lowlands have larger area for cultivation and the lowest altitude compared to mountain areas. Other studies have shown that, mothers living in rural areas are more likely to experience higher infant deaths compared to those residing in urban areas. This can be attributed by poor living standards and or lack of education to most women living in rural areas. Nevertheless, results by Ntimba and Mbago (2005) showed insignificant association when the variable (place of residence) was combined with other socio-demographic factors such as maternal education, wealth index of the household, age at first birth, birth interval and birth order.

### **2.2.8 Household Wealth Index and Infant Mortality**

The wealth index is a method developed by the World Bank aimed at measuring the socioeconomic level of a household in a ranked order. It uses principal component analysis on

the basis of respondents' household assets, amenities, and services. Accordingly, the population is divided into five categories from the poorest fifth to the richest fifth (Mutunga, 2004). In the 2010 TDHS, this variable covered information on ownership of many properties ranging from a television to a bicycle or a car, as well as dwelling characteristics like source of water, sanitation facilities and type of material used in flooring. In low-income countries, because of the difficulty in measuring the income of the households, the wealth index is believed to be a good proxy for measuring the economic status of households (Mutunga, 2004). Generally, several studies have indicated that children born by mothers with poor household index are more likely to experience infant mortality than children born by mothers with rich household index.

### **2.2.9 National Health Insurance**

The National Health Insurance Authority (NHIA) was established under the National Health Insurance Act 2003, Act 650, as a body corporate, with perpetual succession, an Official Seal, that may sue and be sued in its own name. As a body corporate, the Authority in the performance of its functions may acquire and hold movable and immovable property and may enter into a contract or any other transaction. A new law, Act 852 has replaced ACT 650 in October 2012 to consolidate the NHIS, remove administrative bottle necks, introduce transparency, reduce opportunities for corruption and gaming of the system, and make for more effective governance of the schemes. The National Health Insurance Scheme in Ghana is deemed as one of the legacies of the President John Kufuor administration. Seeking the mandate of the people in the 2000 elections, Kufuor promised to abolish what was known at the time as the “cash and carry system” of health delivery. Under this system, patients even those who had been brought into the hospital on emergencies were required to pay money at every point of service delivery. Imagine patients being sent to a hospital with a bleeding accident wound and being asked to pay before a

doctor attends to them. In some cases, lives were lost for the simple reason that friends and relatives were not around to make the required advance payment.

Lack of adequate health infrastructure and knowledge about basic health care among parents pose additional threat to child survival. As a result, many young children die at home before reaching a health clinic or hospital because parents are not informed or do not have easy access to medical services. Health services are inadequate in terms of coverage and quality. Health facilities are few and far between, with 60 per cent of households living more than 30 minutes away from the nearest health facility. These facilities often have limited supplies and drugs, lack suitable sources of water and are staffed by overstretched health workers with insufficient training.

### **2.3 Statistical and Data Mining Techniques**

In the last decade, the extensive effect of classification models on decision making in various fields has attracted much of attention. Classification in the realm of research is the designation of an individual or an item to a set of classes so that the decision making is made based on the characteristic of that individual or the item. Successful classification depends on the two major factors of "how to select the most informative features" and the "classifier method". The widespread in congruency of features in this field has made the selection of a subcategory of the best factors of features more significant, and has given it a more efficient and valuable role in the promotion of the performance of the classification model. Using a set of training patterns, in which the correct classification is known subcategory of classified observations called the training set, the classifier function organizes the classification. Thus, it is expected that proper selection and classification of methods at each stage would lead to a classification model with successful performance. Following the first classification rule presented in 1936 by Fisher in

statistical classification literature (Raudys, 2001), various classification models have been proposed. Among them, is the simple and efficient method for the implementation and understanding of non-parameterized classification was the k-Nearest Neighbor (k-NN) ,which has been well-received. For instance, in order to improve the classification accuracy, Weinberger and Saul(2007) presented a developed algorithm of k-NN. In their proposed model, they used Mahalanobis distance as the criterion for distance determination. A developed hierarchical model of k-NN was introduced by Kubota et al.,(2001). The high capability and sensitivity of this model in the fine discrimination of classes is noteworthy. Zeng et al.(2008) have proposed a modified classification algorithm of k-NN whose underlying algorithm is local average and class statistics. That is, in addition to local information from k-NN of new non-classified data, general information about neighbors in each class is analyzed separately.

Artificial neural network is an efficient approach that in recent years has been considered by researchers as one of the most useful and applicable constructs in artificial intelligence. This is due to its numerous advantages such as being non- parametric (no requirement for any primary assumption on data), self- adaptiveness, ability to be generalized, and having a high capacity in modeling non-linear patterns. This approach is a functional technology that provides the user with the possibility to obtain the best linear combination of features in order to achieve his/her goals including the classification of complex models, estimation of non-linear functions and prediction (Bishop, 1995).

In the medical field, Olmez and Dokur (2012) proposed the use of artificial neural networks algorithm to help classify heart beats. In their proposed model, they first selected the best features using dynamic programming; then, using artificial neural networks, they successfully classified heart beats into seven categories. To classify mortality amongst under five, Rajendra et

al.,(2011) employed artificial neural networks and Fuzzy equivalence classifier. Qiu et al.(2007) presented a model for classification of cervical cancer risk, using artificial neural networks. The findings indicated sensitivity and specificity of 98% and 97%, respectively. Salari et al.,(2013)used artificial neural network methods for prediction of late onset heart failure.Salari et al., (2013) used an integrated medical model based on artificial intelligence approach.

Data mining technique can also be used in bankruptcy prediction as shown by Wilson and Sharda (1994). A major evolution in the studies using financial ratios for bankruptcy prediction was to identify the financial and economic predictors which improve the predictive performance, and two statistical techniques had been used the most: discriminant analysis and logistic regression (Bell, Ribar&Verchio, 1990). Wilson et al., (2003), compared the predictive capability of firm bankruptcy using neural networks and classical multivariate discriminant analysis.

Discriminant analysis is a technique used to build classification schemes so as to assign previously unclassified observation to the appropriate group (Eisenbeis and Avery, 1972). But the underlying assumption for the technique is that the discriminating variable has to be jointly distributed according to a multivariate normal distribution.

Wilson and Sharda (2003) used a number of financial ratios in a multivariate discriminant analysis and contrasted it with the predictive capability of neural network, which is a data mining methodology to show that neural networks performed significantly better than discriminant analysis to predict firm bankruptcy. Statistical techniques have been used to predict the correct placement of a student in the appropriate group as shown by Lin, Huang and Chan (2004). In the same line, Finch and Schneider (2006) conducted a study comparing classification accuracy of

Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Logistic Regression (LR) And Classification And Regression Trees (CART) under a variety of data conditions. Statistical methods for predicting group membership based on a set of measurements have been shown to be very useful in a variety of conditions by Wilson and Handgrave (1995).

Decisions regarding admission to various academic programs, entry into treatment regimens and identification of children at risk for academic failure or behavioral problems were often made with the help of statistical prediction techniques such as Predictive Discriminant Analysis (PDA) or logistic regression (Abedi, 1991; Baird, 1975; Remus & Wong, 1982). PDA has two forms - linear (LDA) and quadratic (QDA). LR is an alternative to PDA and it models the odds of being in one group versus the other as a function of the predictor variable. The CART is a truly non-parametric method because there are no assumptions regarding the underlying distribution from which the subjects are drawn.

Williams, Lee, Fisher and Dickerman (1999) found that both LR and LDA were better at predicting group membership than CART and that QDA performed worse than the other three. But the issue that had not been addressed was the classification accuracy of any of these procedures when one or more of the predictor variables are categorical instead of continuous.

Huberty (1994) recommended using 0 to 1 assignment (dummy coding) and including the variable in the set of predictors when one of the predictors is binary in nature. Johnson and Wichern (2002) suggested that LR might be preferable to LDA when one of the variables is of this type. Finch et al. (2003) conducted a study using Monte Carlo simulations to compare classification accuracy of LDA, QDA, LR and CART and found that QDA approach had a misclassification rate which was never larger than LDA and LR and in many cases it was lower.

When the assumptions of LDA were met, i.e., the data was normally distributed and the covariance matrices of the groups were equal, LDA, LR and QDA had comparable misclassification rates. However, they saw that CART had higher error rates than the other three. The error rates for LDA and LR went up if the data conditions were not met, while QDA and CART's misclassification rates declined when the covariance matrices were not equal. Similar study for comparing cross-validated classification accuracies of predictive discriminant analysis and logistic regression classification models under varying data conditions for a two-group classification problem have been done by Meshbane and Morris (1996).

Among the methods used for solving two-group classification problems, Logistic Regression (LR) And Predictive Discriminant Analysis (PDA) are two of the most popular (Yarnold, Hart and Soltysik, 1994). Several studies have compared the classification accuracy of LR and PDA but the results have been inconsistent. Results of three simulation studies (Baron, 1991; Bayne, Beauchamp, Kane and McCabe, 1983; Crawley, 1979) suggest that LR is more accurate than PDA for non- normal data. However, several researchers (Cleary and Angel, 1984; Dey and Astin, 1993; Knoke, 1982; Krzanowski, 1975; Press and Wilson, 1978) found little or no difference in the accuracy of the two techniques using non-normal data. Findings are also not consistent for degree of group separation.

Bayne et al. (1993) found that larger group separation favored PDA while Crawley (1979) found this condition to favor LR. Sample size is yet another data condition yielding inconsistent results. In a simulation study, Harrell and Lee (1985) found that PDA was more accurate than LR for small samples while in a study by Johnson and Seshia (1992) using real data, LR worked better than PDA for small samples.

Meshbane et al. (1996) proposed a method whereby separate-group as well as total-sample proportions of correct classifications could be compared for the two models using McNemar's test for contrasting correlated proportions and showed that neither theoretical nor data-based considerations were helpful in predicting which of the models would work better. In their study, Hand and Henley (1986) conducted a review of different statistical classification methods used for credit scoring i.e., classifying applicants for credit into 'good' and 'bad' risk classes. The authors examined particular problems arising in the credit scoring context and reviewed the statistical methods, which have been applied. Hand et al.,(2003) mentioned in the study that historically discriminant analysis and linear regression have been most widely used techniques for building score-cards. The first published account of the use of discriminant analysis to produce a scoring system seems to be that of Durand (1941) who showed that the method could produce good predictions of credit replacement.

Myers and Forgy (1963) had compared discriminant analysis and regression analysis for credit scoring and Grablowsky and Talley (1981) compared linear discriminant analysis and probit analysis for the same purpose. Wiginton (1980) gave one of the first published accounts of logistic regression applied to credit scoring in comparison to discriminant analysis and concluded that logistic regression gave a superior result.

Rosenberg and Gleit (1994) described several applications of neural networks to corporate credit decisions and fraud detection and Davis, Edelman and Gammerman (1992) compared such methods with alternative classifiers. Non-parametric methods, especially nearest neighbor methods, have been explored for credit scoring applications by Chatterjee and Barcun (1970) and Hand (1986). In addition to the mentioned methods, Hand et al.(1986), also considered mathematical programming methods, recursive partitioning, expert systems and time varying

methods, summarized the various methods in their study, assessed the relative strengths and weaknesses of the methods and have drawn the conclusion that there is no overall 'best' method. What is best depends on the details of the problem: on the data structure, the characteristics used the extent to which it is possible to separate the classes by using those characteristics and the objective of the classification (overall misclassification rate, cost-weighted misclassification rate, bad risk among those accepted, some measure of profitability, etc.). If the classes are not well separated, then  $P_r$  (good risk (characteristic vector)) is a flat function, so that the decision separating the classes cannot be accurately estimated. In such circumstances, highly flexible methods such as neural networks and nearest neighbor methods are vulnerable to over fitting the design data and considerable smoothing must be used. Nearest neighbor methods are effective with regard to the speed of classification. Neural networks are well suited to situations where there is a poor understanding of the data structure. If there is a good understanding of data structure and the problem, methods which make use of this understanding, such as regression, nearest neighbor and tree-based methods are expected to perform better. The authors infer that in credit scoring, since people have been constructing score-cards on similar data for decades, there is solid understanding and hence, neural networks have not been adopted as a regular production system. Henley and Hand (1996) have also studied the application of k-nearest-neighbor (k-NN) method, a standard technique in pattern recognition and nonparametric statistics, as a credit scoring techniques for assessing the credit worthiness of consumer loan applicants. The k-NN method is a standard non-parametric technique used for probability density function estimation and classification and was originally proposed by Fix and Hodges (1952) and Cover and Hart (1967).

Henley et al.,(1996) proposed this study to provide a practical classification model that can improve on traditional credit scoring techniques. The authors proposed an adjusted version of the Euclidean distance metric which attempted to incorporate knowledge of class separation contained in data. To assess the potential of this method, the authors, drew a comparison k-NN with linear and logistic regression and decision trees and graphs and showed that the k-NN method with adjusted Euclidean metrics can give slightly improved prediction of consumer credit risk than the traditional techniques, achieving the lowest expected bad risk rate. It has been observed that most cases that are misclassified by one method can be correctly predicted by other approaches (Tam and Kiang, 1992). A study on comparative analysis of ID3 and neural networks conducted by Dieterrich, Hild and Bakiri (1995) also had similar observations.

Breiman (1996) studied the instability of different predictors and concluded that neural networks, classification trees and subset selection in linear regression were unstable while the k-th nearest neighbor method was stable.

A study by Anderson (1984) revealed that under the assumptions of multivariate normal distributions with known parameters and equal covariance matrices in the classes, linear classifiers provide optimal classification. In their study, Asparoukhov et al., (2003) concluded that the traditional statistical classifiers were not well able to cope with small sample binary data but the non-traditional (MLP, LVQ, MIP) classifiers did much better under those circumstances. Another interesting study for comparison between neural networks and logistic regression for predicting patronage behavior towards web and traditional stores has been done by Chiang, Zhang and Zhou (2006). Different kinds of empirical studies for predicting customer preference for online shopping have been done (Degeratu, Rangaswamy and Wu, 2000; Bellman, Lohse and Johnson, 1999; Kwak, Fox and Zinkhan, 2002). Chiang et al.(2006), developed neural network

models which are known for their known capability of modeling non-compensatory decision processes and tried to find out whether non-compensatory choice models using neural network perform better than logit choice models in predicting consumer's channel choice between web and traditional stores. The authors showed that for most of the selected products, neural networks significantly outperform logistic regression models in terms of predictive power. Studies by Fadlalla and Lin (2001), Hung, Liang and Liu (1996) and West, Brockett and Golden (1997) also show that in most of the applications where neural networks have been used to model business problems in support of finance and marketing decision-making, neural networks have outperformed traditional compensatory models such as discriminant and regression analysis. Kim et al.,(2003) used artificial neural networks (Riedmiller, 1994) guided by genetic algorithms (Goldberg, 1989) to develop their predictive model.

In a similar study, Chu and Widjaja (1994) showed that neural networks using a back-propagation based forecasting prototype can be effectively used as a forecasting tool. A key study with respect to comparative assessment of classification methods has been done by Kiang (2003). In this study Kiang has considered data mining classification techniques viz.neural networks and decision tree models and three statistical methods - Linear Discriminant Analysis (LDA), Logistic Regression Analysis And K-Nearest-Neighbor (Knn) models, and used synthetic data to perform a controlled experiment in which the data characteristics are systematically altered to introduce imperfections such as nonlinearity, multicollinearity, unequal covariance, etc. The study was performed to investigate how these different classification methods performed when certain assumptions about the data characteristics were violated and the author showed that data characteristics considerably impacted the classification performance of the methods. Also, the study conducted by Shavlik, Mooney and Towell (1991) emphasized

by empirically analyzing the effects of three factors on the performance of two AI methods, neural networks and ID3. The three factors considered were size of training data, imperfect training examples and encoding of the desired outputs. Shavlik et al.,(1991) showed that neural networks performed well with small sizes of training data but they did not emphasize much on the distribution of the data instances.



## **CHAPTER THREE**

### **METHODOLOGY**

#### **3.1 Introduction**

This chapter provides an outline on how the research methodology was used in the investigation of the proposed research problem with specific reference to the description of the data source, method of analysis and the type of statistical tools employed. The data mining software SPSS Version 20 was used for building the models.

Three different modeling techniques were considered in this research: Logistic regression statistical method, nearest neighbor classification and neural network. These are data mining methods. The data used were taken from the 2008 Ghana Demographic and Health Survey (GDHS).

#### **3.2.0 Data Source**

This study was undertaken using the 2008 Ghana Demographic and Health Survey (GDHS). The 2008 GDHS was carried out by the Ghana Statistical Service and the Ghana Health Service (GHS), ICF Macro, an ICF International Company, provided technical support for the survey through the MEASURE DHS programme. The survey was designed to obtain information on the demographic and socioeconomic status of respondents as well as monitor the health situation in Ghana. The GDHS used a representative sample of 12,000 households nationwide including all the ten regions as well as rural and urban areas. Interviews were requested from all women age

15-49 and all men age 15-59 in the households that were sampled. Both usual residents and visitors who had spent the previous night in the household were interviewed.

Questionnaires were translated from English into the five major languages of Ghana (Akan, Nzema, Ewe, Ga, and Dagbani). Three separate questionnaires were utilized in each household which include the household questionnaire, which obtained information on household water source and sanitation facilities; a women's questionnaire, including information on childhood illness, and a men's questionnaire, which was not used in this study. This study worked with the children data set which was downloaded online from the Measure DHS website, which incorporated information on household as well as information on child mortality reported by mothers for all children under age five.

Dependent Variable: "Is child alive"

The possible explanatory variables to consider are:

Type of place of residence of family, Highest educational level of mother, Source of drinking water, type of toilet facility used by family, Wealth index of family, alcohol consumption of mother, Whether mother is covered by health insurance, Marital status of mother, Sex of child and age of mother at child birth.

### **3.3 Logistic Regression**

Logistic regression is a mathematical modelling approach that can be used to describe the relationship of several independent variables to a dichotomous dependent variable. Logistic regression allows the prediction of discrete variables by a mix of continuous and discrete predictors. Other modelling approaches are possible also, but logistic regression is by far the

most popular modelling procedure used to analyse epidemiologic data when the illness measure is dichotomous. It addresses the same questions that discriminant function analysis and multiple regression do but with no distributional assumptions on the predictors thus the predictors do not have to be normally distributed, linearly related or have equal variance in each group.

### **3.3.1.1 What Logistic Regression Predicts**

The value produced by logistic regression is a probability value between 0.0 and 1.0. If the probability for group membership in the modeled category is above some cut point (the default is 0.50), the subject is predicted to be a member of the modeled group. If the probability is below the cut point, the subject is predicted to be a member of the other group. For any given case, logistic regression computes the probability that a case with a particular set of values for the independent variable is a member of the modeled category.

### **3.3.1.2 Level of Measurement Requirements**

Logistic regression analysis requires that the dependent variable be dichotomous. Logistic regression analysis requires that the independent variables be metric or dichotomous. If an independent variable is of a nominal level and not dichotomous, the logistic regression procedure in SPSS has an option to dummy code the variable. If an independent variable is ordinal, the usual caution is attached.

### **3.3.1.3 Assumptions**

Logistic regression does not make any assumptions of normality, linearity, and homogeneity of variance for the independent variables. Because it does not impose these requirements, it is preferred to discriminant analysis when the data does not satisfy these assumptions.

#### **3.3.1.4 Sample Size Requirements**

The minimum number of cases per independent variable is 10, using a guideline provided by Hosmer and Lemeshow (2000), authors of *Applied Logistic Regression*, one of the main resources for Logistic Regression. For preferred case-to-variable ratios, 20 to 1 is used for simultaneous and hierarchical logistic regression and 50 to 1 for stepwise logistic regression.

#### **3.3.2 Methods of Including Variables**

There are three methods available for including variables in the regression equation: the simultaneous method in which all independents are included at the same time. The hierarchical method in which control variables are entered in the analysis before the predictors whose effects are the primary concern of the analysis. The stepwise method (forward conditional in SPSS) in which variables are selected in the order in which they maximize the statistically significant contribution to the model is employed. For all methods, the contribution to the model is measured by model chi-square, which is a statistical measure of the fit between the dependent and independent variables, like  $R^2$  in multiple regression analysis.

#### **3.3.3 Maximum Likelihood (ML) Estimation**

Maximum likelihood (ML) estimation is an alternative approach developed by statisticians for estimating the parameters in a mathematical model. Until the availability of computer software for ML estimation, the method used to estimate the parameters of a logistic model was discriminant function analysis. There are actually two alternative ML approaches that can be used to estimate the parameters in a logistic model. These are called the unconditional method and the conditional method. These two methods require different computer programs. Three of the most widely available computer packages for unconditional ML estimation of the logistic model are SAS, SPSS, and Stata. Programs for conditional ML estimation are available in all

three packages, but some are restricted to special cases. In making the choice between unconditional and conditional ML approaches, the researcher needs to consider the number of parameters in the model relative to the total number of subjects under study. In general, unconditional ML estimation is preferred if the number of parameters in the model is small relative to the number of subjects. In contrast, conditional ML estimation is preferred if the number of parameters in the model is large relative to the number of subjects. To describe the ML procedure, the introduction of the likelihood function,  $L$  is made. The likelihood function can alternatively be denoted as  $L(\theta)$ . The likelihood function  $L$  or  $L(\theta)$  represents the joint probability or likelihood of observing the data that has been collected. The term “joint probability” means a probability that combines the contributions of all the subjects in the study. In matrix terminology, the collection  $\theta$  is referred to as a vector; its components are the individual parameters being estimated in the model, denoted here as  $\theta_1, \theta_2, \dots, \theta_q$ , where  $q$  is the number of individual components.

As described earlier, if the model is logistic, there are two alternative types of computer programs to choose from, an unconditional versus a conditional program. These programs use different likelihood functions, namely,  $L_u$  for the unconditional method and  $L_c$  for the conditional method. The unconditional formula is given first, and directly describes the joint probability of the study data as the product of the joint probability for the cases (diseased persons) and the joint probability for the non-cases (non-diseased persons).

$$L_u = \prod_{l=1}^{m_1} P(X_l) \prod_{l=m_1+1}^n [1 - P(X_l)] \quad (3.1)$$

The probability of obtaining the data for the  $i$ th case is given by  $P(X_i)$

The probability of the data for the  $i$  th non case is given by  $1 - P(X_i)$

$$P(X) = \text{Logistic model} = \frac{1}{e^{-(\alpha + \sum \beta_i X_i)}} \quad (3.2)$$

Where  $X_i$ 's are the independent variables,  $\alpha$  and the  $\beta_i$  are constant terms representing unknown parameters. When the logistic model formula involving the parameters is substituted into the likelihood expression 3.1, the formula shown in 3.3 is obtained after a certain amount of algebra is done.

$$L_u = \frac{\prod_{i=1}^n \exp(\alpha + \sum_{i=1}^k \beta_i X_{il})}{\prod_{i=1}^n \left[ 1 + \exp(\alpha + \sum_{i=1}^k \beta_i X_{il}) \right]} \quad (3.3)$$

The conditional likelihood formula ( $L_c$ ) reflects the probability of the observed data configuration relative to the probability of all possible configurations of the given data. To understand this, the researcher describes the observed data configuration as a collection of  $m_1$  cases and  $n - m_1$  non cases. The researcher denotes the cases by the  $\mathbf{X}$  vectors  $\mathbf{X}_1, \mathbf{X}_2$ , and so on through  $\mathbf{X}_{m_1}$  and the non-cases by  $X_{m_1+1}, X_{m_1+2}$ , through  $X_n$ .

$$L_c = \frac{\text{Pr(Observed data)}}{\text{Pr(all possible configurations)}}$$

$m_1$  cases :  $(X_1, X_2, \dots, X_{m_1})$

$n - m_1$  Non-cases:  $(X_{m_1+1}, X_{m_1+2}, \dots, X_n)$

$L_c = \text{Pr}(\text{first } m_1 X_1 \text{ are cases} \mid \text{all possible configurations of X's})$

The configuration in 3.3 assumes that the observed data has been rearranged so that the  $m_1$  cases are listed first and are then followed in listing by the  $(n - m_1)$  non-cases. Using this configuration, the conditional likelihood function gives the probability that the first  $m_1$  of the observations actually go with the cases, given all possible configurations of  $3.3n$  observations into a set of  $m_1$  cases and a set of  $n - m_1$  non-cases. The term configuration here refers to one of the possible ways that the observed set of  $X$  vectors can be partitioned into  $m_1$  cases and  $n - m_1$  non-cases.

Possible Configurations = combinations of  $n$  things taken  $m_1$

at a time =  $C_{m_1}^n$

The formula for the conditional likelihood is then given by the expression 3.4.

$$L_c = \frac{\prod_{l=1}^{m_1} P(X_l) \prod_{l=m_1+1}^n [1 - P(X_l)]}{\sum_u \left\{ \prod_{l=1}^{m_1} P(X_{ul}) \prod_{i=m_1+1}^n [1 - P(X_{ul})] \right\}} \quad (3.4)$$

When the logistic model formula involving the parameters is substituted into the conditional likelihood expression above, the resulting formula shown in 3.4 is obtained. This formula is not the same as the unconditional formula shown in 3.1. Moreover, in the conditional formula, the intercept parameter  $\alpha$  has dropped out of the likelihood.

The removal of the intercept  $\alpha$  from the conditional likelihood is important because it means that when a conditional ML program is used, estimates are obtained only for the  $\beta_i$  coefficients in the model and not for  $\alpha$ . Because the usual focus of a logistic regression analysis is to estimate an

odds ratio, which involves the  $\beta$ 's and not  $\alpha$ , one does not usually care about estimating and, therefore, consider  $\alpha$  to be a nuisance parameter.

The conditional likelihood formula without the intercept is shown in 3.5;

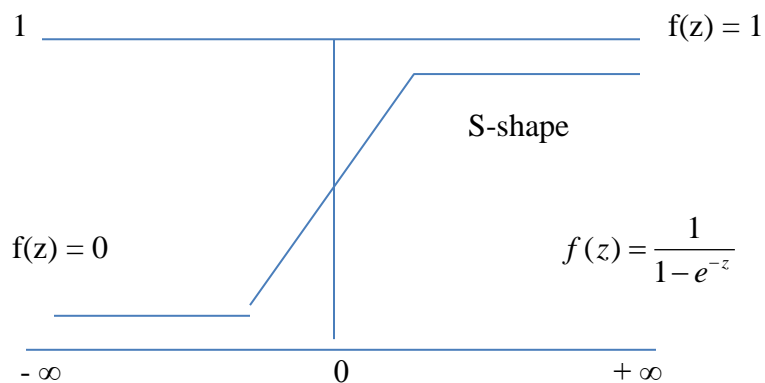
$$L_c = \frac{\prod_{l=1}^{m_i} \exp\left(\sum_{i=1}^k \beta_i X_{li}\right)}{\sum_u \left[ \prod_{l=1}^{m_i} \exp\left(\sum_{i=1}^k \beta_i X_{lui}\right) \right]} \quad (3.5)$$

### 3.3.4 Overall Test Of Relationship

The overall test of relationship among the independent variables and groups defined by the dependent is based on the reduction in the likelihood values for a model, which does not contain any independent variables and the model that contains the independent variables. This difference in likelihood follows a chi-square distribution, and is referred to as the model chi-square. The significance test for the model chi-square is the statistical evidence of the presence of a relationship between the dependent variable and the combination of the independent variables.

### 3.3.5 The Logistic Function

The logistic function called  $f(z)$ , describes the mathematical form on which the logistic model is based. This function, called  $f(z)$ , is given by 1 over 1 plus  $e$  to the minus  $z$ .



**Figure 3.1: Graph showing the non-linearity of the logistic regression function**

The fact that the logistic function  $f(z)$  ranges between 0 and 1 is the primary reason the logistic model is so popular. The model is designed to describe a probability, which is always some number between 0 and 1. In epidemiologic terms, such a probability gives the risk of an individual getting a disease.

### 3.3.6 The Logit Transformation

The logit transformation, denoted as  $\text{logit } P(X)$ , is given by the natural log (i.e., to the base  $e$ ) of the quantity  $P(X)$  divided by one minus  $P(X)$ , where  $P(X)$  denotes the logistic model as previously defined. This transformation allows us to compute a number, called  $\text{logit } P(X)$ , for an individual with independent variables given by  $X$ . We do so by:

- (i) computing  $P(X)$  and
- (ii) 1 minus  $P(X)$  separately, then
- (iii) dividing one by the other, and finally
- (iv) taking the natural log of the ratio.

$$\text{Logit } P(X) = \ln \left[ \frac{P(X)}{1 - P(X)} \right]$$

44

$$\text{where } P(X) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i x_i)}} \quad (3.6)$$

We can write  $1 - P(X)$  as;

$$1 - P(X) = 1 - \frac{1}{1 - e^{-(\alpha + \sum \beta_i X_i)}}$$

$$1 - P(X) = \frac{e^{-(\alpha + \sum \beta_i X_i)}}{1 + e^{-(\alpha + \sum \beta_i X_i)}} \quad (3.7)$$

If we divide  $P(X)$  by  $1 - P(X)$ , then the denominators cancel out and we obtain the result below;

$$\frac{P(X)}{1 - P(X)} = \frac{\frac{1}{1 - e^{-(\alpha + \sum \beta_i X_i)}}}{\frac{e^{-(\alpha + \sum \beta_i X_i)}}{1 + e^{-(\alpha + \sum \beta_i X_i)}}}$$

$$\frac{P(X)}{1 - P(X)} = e^{(\alpha + \sum \beta_i X_i)} \quad (3.8)$$

We go on to compute the natural log of the formula just derived to obtain:

$$\ln_e \left[ \frac{P(X)}{1 - P(X)} \right] = \ln_e \left[ e^{(\alpha + \sum \beta_i X_i)} \right]$$

$$= \alpha + \sum \beta_i X_i$$

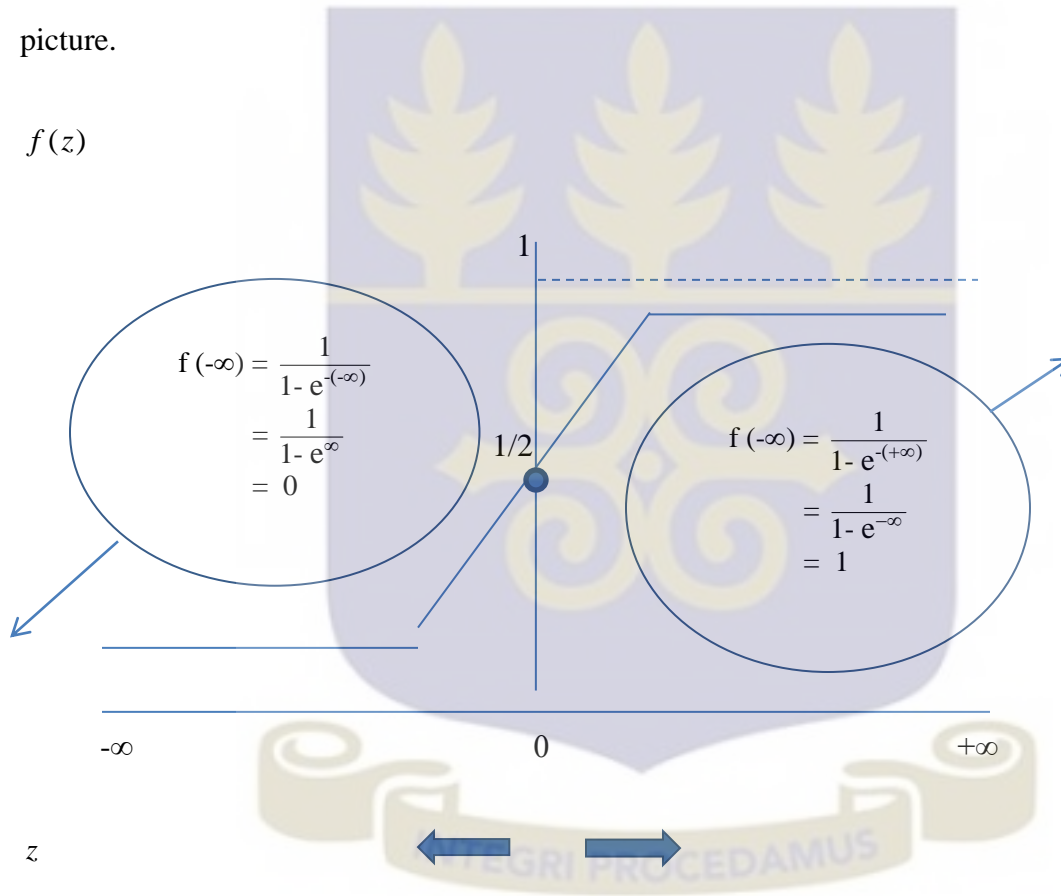
Hence in logit form,

$$\text{Logit } P(X) = \alpha + \sum \beta_i X_i \quad (3.9)$$

### 3.3.7 The Logistic Model

The logistic model, therefore, is set up to ensure that whatever estimate of risk is obtained, it will always be some number between 0 and 1. Thus, for the logistic model, one can never get a risk

estimate either above 1 or below 0. This is not always true for other possible models, which is why the logistic model is often the first choice when a probability is to be estimated. Another reason why the logistic model is popular derives from the shape of the logistic function. As shown in the graph (Figure 3.1), if one starts at  $z = -\infty$  and moves to the right, then as  $z$  increases, the value of  $f(z)$  hovers close to zero for a while, then starts to increase dramatically toward 1, and finally levels off around 1 as  $z$  increases toward  $+\infty$ . The result is an elongated, S-shaped picture.



**Figure 3.2:** A plot of  $f(z)$  as  $z$  takes values from  $-\infty$  to  $+\infty$

To obtain the logistic model from the logistic function, we write  $z$  as the linear sum  $\alpha$  plus  $\beta_1$  times  $X_1$  plus  $\beta_2$  times  $X_2$ , and so on to  $\beta_k$  times  $X_k$ , where the  $X$ 's are independent variables of interest and  $\alpha$  and the  $\beta_i$  are constant terms representing unknown parameters.

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k = \alpha + \sum \beta_i X_i \quad (3.10)$$

In essence, then,  $z$  is an index that combines the  $X$ 's. Then, one substitutes the linear sum expression for  $z$  in the right-hand side of the formula for  $f(z)$  to get the expression for  $f(z)$  in an epidemiologic context as follows.

$$f(z) = \frac{1}{1 + e^{-z}}$$

$$f(z) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}} \quad (3.11)$$

### 3.3.8 HosmerLemeshow (H-L) Goodness Of Fit Statistic

There are, in fact, a number of different measures of goodness of fit for logistic regression models. One of these measures is the Hosmer-Lemeshow test of goodness of fit. This is similar to a Chi Square test, and indicates the extent to which the model provides better fit than a null model with no predictors, or, in a different interpretation, how well the model fits the data, as in log-linear modelling. If chi-square goodness of fit is not significant, then the model has adequate fit. By the same token, if the test is significant, the model does not adequately fit the data. The Hosmer and Lemeshow's (H-L) goodness of fit test divides subjects into deciles based on predicted probabilities and then computes a chi-square from observed and expected frequencies. Then a probability ( $p$ ) value is computed from the chi-square distribution to test the fit of the logistic model. If the H-L goodness-of-fit test statistic is greater than 0.05, as demanded for well-fitting models, we fail to reject the null hypothesis that there is no difference between observed and model-predicted values, implying that the model's estimates fit the data at an acceptable level. That is, well-fitting models show non-significance on the goodness-of-fit test, indicating model prediction that is not significantly different from observed values.

A disadvantage of this goodness of fit measure is that it is a significance test, with all the limitations this entails. Like other significant tests it only tells us whether the model fits or not, and does not state anything about the extent of the fit. Similarly, like other significance tests, it is strongly influenced by the sample size (sample size and effect size both determine significance), and in large samples, a very small difference will lead to significance. As the sample size gets large, the H-L statistic can find smaller and smaller differences between observed and model-predicted values to be significant. Small sample sizes are also problematic, however, as, being a chi Square test one cannot have too many groups (more than 10%) with predicted frequencies of less than five.

The test statistic is obtained by applying a chi-square test on a  $2 \times g$  contingency table. The contingency table is constructed by cross-classifying the dichotomous dependent variable with a grouping variable (with  $g$  groups) in which groups are formed by partitioning the predicted probabilities using the percentiles of the predicted event probability. In the calculation, approximately 10 groups are used ( $g = 10$ ). The corresponding groups are often referred to as the “deciles of risk” (Hosmer and Lemeshow, 1989).

If the values of independent variables for observation  $i$  and  $i'$  are the same, observation  $i$  and  $i'$  are said to be in the same block. When one or more blocks occur within the same decile, the blocks are assigned to this same group. Moreover, observations in the same block are not divided when they are placed into groups. This strategy may result in fewer than 10 groups (that is,  $g \leq 10$ ) and consequently, fewer degrees of freedom.

Suppose that there are  $Q$  blocks, and the  $q$ th block has  $m_q$  number of observations,  $q = 1, \dots, Q$ . Moreover, suppose that the  $k$ th group ( $k = 1, \dots, g$ ) is composed of the  $q_1$ th, ...,  $q_k$ th blocks of

observations. Then the total number of observations in the  $k^{\text{th}}$  group is  $S_k = \sum_{q_1}^{q_k} m_j$ . The total observed frequency of events (that is,  $Y = 1$ ) in the  $k^{\text{th}}$  group, call it  $O_{1k}$ , is the total number of observations in the  $k^{\text{th}}$  group with  $Y = 1$ . Let  $E_{1k}$  be the total expected frequency of the event (that is,  $Y = 1$ ) in the  $k^{\text{th}}$  group; then  $E_{1k}$  is given by  $E_{1k} = S_k \xi_k$ , where  $\xi_k$  is the average predicted event probability for the  $k^{\text{th}}$  group.

$$\xi_k = \sum_{q_1}^{q_k} m_j \hat{\pi}_j / S_k \quad (3.12)$$

Where  $\hat{\pi}_j$  is the probability of  $Y$  for the  $j^{\text{th}}$  case

The Hosmer-Lemeshow goodness-of-fit statistic is computed as

$$\chi_{hl}^2 = \sum_{k=1}^g \frac{(O_{1k} - E_{1k})^2}{E_{1k}(1 - \xi_k)} \quad (3.13)$$

The  $p$  value is given by  $Pr(\chi^2 = \chi_{hl}^2)$  where  $\chi^2$  is the chi-square statistic distributed with degrees of freedom ( $g = 2$ ).

### 3.11 Wald Statistic

The Wald statistic is calculated for the variables in the model to determine whether a variable should be removed. If the  $i^{\text{th}}$  variable is not categorical, the Wald statistic is defined by

$$Wald_i = \frac{\hat{\beta}_i}{\hat{\sigma}_{\hat{\beta}_i}^2} \quad (3.14)$$

If it is a categorical variable, the Wald statistic is computed as follows:

Let  $\hat{\beta}_i$  be the vector of maximum likelihood estimates associated with the  $(m-1)$  dummy variables, and  $C$  the asymptotic covariance matrix for  $\hat{\beta}_i$ . The Wald statistic is

$$Wald_i = \hat{\beta}_i C^{-1} \hat{\beta}_i \quad (3.15)$$

The asymptotic distribution of the Wald statistic is chi-square with degrees of freedom equal to the number of parameters estimated.

### 3.3.9 Odds Ratio

#### 3.3.9.1 Definition of Odds Ratio

An odds ratio (OR) is a measure of association between an exposure and an outcome. The OR represents the odds that an outcome will occur given a particular exposure, compared to the odds of the outcome occurring in the absence of that exposure. Odds ratios are most commonly used in case-control studies, however they can also be used in cross-sectional and cohort study designs as well (with some modifications and/or assumptions). When a logistic regression is calculated, the regression coefficient ( $b_1$ ) is the estimated increase in the log odds of the outcome per unit increase in the value of the exposure. In other words, the exponential function of the regression coefficient ( $e^{b_1}$ ) is the odds ratio associated with a one-unit increase in the exposure.

In its simplest form, an odds is the ratio of the probability that some event will occur over the probability that the same event will not occur. The formula for an odds is, therefore, of the form  $P$  divided by  $1 - P$ , where  $P$  denotes the probability of the event of interest.

$$Odds = \frac{1}{1 - p} \quad (3.16)$$

Odds ratios are used to compare the relative odds of the occurrence of the outcome of interest (e.g. disease or disorder), given exposure to the variable of interest (e.g. health characteristic, aspect of medical history). The odds ratio can also be used to determine whether a particular exposure is a risk factor for a particular outcome, and to compare the magnitude of various risk factors for that outcome.

- OR=1 Exposure does not affect odds of outcome
- OR>1 Exposure associated with higher odds of outcome
- OR<1 Exposure associated with lower odds of outcome

The 95% Confidence Interval (CI) is used to estimate the precision of the OR. A large CI indicates a low level of precision of the OR, whereas a small CI indicates a higher precision of the OR. It is important to note however, that unlike the p value, the 95% CI does not report a measure's statistical significance. In practice, the 95% CI is often used as a proxy for the presence of statistical significance if it does not overlap the null value (e.g. OR=1). Nevertheless, it would be inappropriate to interpret an OR with 95% CI that spans the null value as indicating evidence for lack of association between the exposure and outcome.

### 3.3.9.2 Derivation of the Odds Ratio Formulae

Any odds ratio, by definition, is a ratio of two odds, written here as **odds<sub>1</sub>** divided by **odds<sub>2</sub>**, in which the subscripts indicate two individuals or two groups of individuals being compared.

$$OR = \frac{Odds_1}{Odds_2}$$

Let  $\mathbf{X}_1$  denote the collection of  $X$ 's that specify group 1 and let  $\mathbf{X}_0$  denote the collection of  $X$ 's that specify group 0.

$$X_1 = (X_{11}, X_{12}, \dots, X_{1k})$$

$$X_2 = (X_{21}, X_{22}, \dots, X_{2k})$$

Given a logistic model of the general form  $P(X)$ ,

$$P(X) = \frac{1}{1 + e^{-(\alpha + \sum \beta X_i)}} \quad (3.17)$$

The Odds for group 1 is given by;  $Odds_1 = \frac{P(X_1)}{1 - P(X_1)}$

The Odds for group 2 is given by;  $Odds_2 = \frac{P(X_2)}{1 - P(X_2)}$

The ratio of the Odds for  $X_1$  to the ratio of  $X_2$  is defined as the Risk Odds ratio (ROR), which is represented mathematically by the expression below.

$$ROR = \frac{\frac{P(X_1)}{1 - P(X_1)}}{\frac{P(X_2)}{1 - P(X_2)}}$$

$$\frac{P(X_1)}{1 - P(X_1)} = e^{(\alpha + \sum \beta_i X_{1i})} \text{ and } \frac{P(X_2)}{1 - P(X_2)} = e^{(\alpha + \sum \beta_i X_{2i})}$$

$$ROR_{X_1, X_2} = \frac{Odds_1}{Odds_2} = \frac{e^{(\alpha + \sum \beta_i X_{1i})}}{e^{(\alpha + \sum \beta_i X_{2i})}}$$

$$= e^{(\alpha + \sum \beta_i X_{1i}) - (\alpha + \sum \beta_i X_{2i})}$$

$$= e^{\sum \beta_i (X_{1i} - X_{2i})}$$

The formulae is generalized for the ROR which becomes;

$$ROR_{X_1, X_2} = e^{\sum_{i=1}^k \beta_i (X_{1i} - X_{2i})} \quad (3.18)$$

A general exponential formula is obtained for the risk odds ratio from a logistic model comparing any two groups of individuals, as specified in terms of  $X_1$  and  $X_0$ . Note that the formula involves the  $\beta_i$ 's but not  $\alpha$ .

### 3.3.10 The Likelihood Ratio (LR) Test

The likelihood-ratio test uses the ratio of the maximized value of the likelihood function for the full model ( $L_1$ ) over the maximized value of the likelihood function for the simpler model ( $L_0$ ). The full model has all the parameters of interest in it. The simpler model is said to be a nested, reduced model, where an independent variable is dropped from the overall model. The likelihood-ratio test tests if the logistic regression coefficient for the dropped variable can be treated as zero, thereby justifying the dropping of the variable from the model. A non-significant likelihood-ratio test indicates no difference between the full model and the reduced model, hence justifying dropping the given variable so as to have a more parsimonious model that works just as well. The likelihood-ratio test statistic equals:

$$-2 \log(L_0/L_1) = -2[\text{Log}(L_0) - \text{Log}(L_1)] = -2(\ell_0 - \ell_1) \quad (3.19)$$

Where  $\ell_0 = \text{Log}(L_0)$  and  $\ell_1 = \text{Log}(L_1)$

### 3.3.11 Regression Coefficients

The regression coefficients are the coefficients  $b_1, b_2, \dots, b_k$  of the regression equation:

$$\text{Logit}(p) = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + \dots + b_k X_k \quad (3.20)$$

An independent variable with a regression coefficient which is not significant can be removed from the regression model. If  $P < 0.05$ , then the variable contributes significantly to the prediction of the outcome variable. The logistic regression coefficients show the change

(increase when  $b_i > 0$ , decreases when  $b_i < 0$ ) in the predicted log odds of having the characteristic of interest for a one- unit change in the independent variables.

Given the model,  $\text{logit } P(X) = \alpha + \sum \beta_i X_i$  refers to the effect of  $X_i$  on the log odds  $D = 1$  controlling the other  $X_i$ . This provides basic interpretation for the magnitude of  $\beta_i$ . For instance  $e^{\beta_i}$  is the multiplicative effect on the odds of  $s_1$  – unit increase in  $x_1$ , at fixed level other  $x_j$ . The sign of  $\beta_i$  determines whether  $P(X)$  is increasing or decreasing as  $x$  increases. The rate of climb or decent increases as  $|\beta|$  increases; as  $\beta \rightarrow 0$  the curve straightens a horizontal straight line when  $\beta = 0$ , is independent of  $X$ . For quantitative  $X_i$  with  $\beta > 0$ , the curve for  $P(X)$  approaches 1 at the same rate that it approaches 0.

### 3.3.12 P-Value

P-value is associated with a test statistic. It is "the probability, if the test statistic really were distributed as it would be under the null hypothesis, of observing a test statistic [as extreme as, or more extreme than] the one actually observed". The smaller the  $P$  value, the more strongly the test rejects the null hypothesis, that is, the hypothesis being tested. A p-value of .05 or less rejects the null hypothesis "at the 5% level" that is, the statistical assumptions used imply that only 5% of the time would the supposed statistical process produce a finding this extreme if the null hypothesis were true. Significance levels of 5% and 10% are common figures to which p-values are compared.

### 3.3.13 Overall Model Fit

The null model -2 log likelihood is given by  $-2\ell_0$  where  $L_0$  is the likelihood of obtaining the observations if the independent variables had no effect on the outcome. The full model -2 Log

likelihood is given by  $-2 \ln(L)$  where  $L$  is the likelihood of obtaining the observations with all independent variables incorporated in the model. The difference of these two yields a Chi-square statistic which is a measure of how well the independent variables affect the outcome or dependent variable. If the  $p$ -value for the overall model fit statistic is less than the conventional 0.05 then there is evidence that at least one of the independent variables contributes to the prediction of the outcome.

#### **3.3.14 Study Assumptions**

All responses and information provided by the respondents are assumed to be accurate and true characteristics of the study area.

#### **3.3.15 Omnibus Tests of Model Coefficients**

The omnibus test of model coefficients was used to establish the presence of a relationship between the dependent variable and combination of independent variables. It was based on the statistical significance of the model chi-square at step 1 after the independent variables have been added to the analysis.

#### **3.3.16 Logistic Regression Table**

The results in the final logistic regression model, which includes estimated model parameters, their standard errors, odds ratio, Wald tests, and associated  $P$ -values were presented in a table. The parameter estimates were the estimated coefficients of the fitted logistic regression model. The logistic regression equation was also obtained.

### **3.3.17 HosmerAndLemeshow Goodness Of Fit Test**

The Hosmer and Lemeshow Goodness of Fit Test was used to determine if the hypothesized model fits the data set used in predicting mortality likelihood among under five children in Ghana.

### **3.3.18 Model Summary**

-2 Log likelihood statistic was used to measure how poorly the model predicts the decisions, the smaller the statistic the better the model. Also Cox & Snell R Square and Nagelkerke R Square which determines how much of the variance in the dependent variable is explained by the model was computed.

A more useful measure to assess the utility of a logistic regression model is classification accuracy, which compares predicted group membership based on the logistic model to the actual, known group membership, which is the value for the dependent variable. A Classification Table was used to present the classification accuracy.

### **3.3.19 Explanation Of Variables In The Model**

The probability of the Wald statistic was computed for each variable in the model to determine whether a variable should be removed or maintained to determine whether they are significant in predicting mortality in children under five years. The corresponding Odds ratio, which is the likelihood that an event will occur given a particular exposure was calculated for each variable in the model. The issue of multicollinearity in the logistic regression model was detected by examining the standard errors for the  $\beta$  coefficients. A standard error larger than 2.0 indicates numerical problems, such as multicollinearity among the independent variables.

### 3.4 Neural Network

A neural network is a software (or hardware) simulation of a biological brain (sometimes called Artificial Neural Network or 'ANN'). The purpose of a neural network is to learn to recognize patterns in a given data set. In the human brain, a typical neuron collects signals from others through a host of fine structures called *dendrites*. The neuron sends out spikes of electrical activity through a long thin strand known as an *axon*, which splits into thousands of branches. At the end of each branch, a structure called a *synapse* converts the activity from the axon into electrical signals that inhibit or excite activity in the connected neurons. When a neuron receives excitatory input that is sufficiently large compared with its inhibitory input, it sends a spike of electrical activity down its axon. Learning occurs by changing the effectiveness of the synapses so that the influence of one neuron on another changes.

These neural networks may be built by typically programming to emulate the essential features of neurons and their interconnections. However, because the knowledge of neurons is incomplete and computing power is limited, the models are necessarily gross idealizations of real networks of neurons. An important application of neural network is pattern recognition, which can be implemented using a feed-forward neural network that has been trained accordingly. During training, the network is trained to associate outputs with input patterns. When the network is used, it identifies the input pattern and tries to output the associated output pattern. The power of neural network comes to life when a pattern that has no output associated with it, is given as an input. In this case, the network gives the output that corresponds to a taught input pattern that is least different from the given input pattern.

Neural networks are capable of modeling extremely complex, typically non-linear functions. Each neuron has a certain number of inputs, each of which has a weight assigned to it. The

weight is an indication of the importance of the incoming signal for that input. These weighted inputs are added together and if they exceed a pre-set threshold value, the neuron fires. The input value received from a neuron is calculated by summing the weighted input values from its input links. An activation function takes the neuron input value and produces a value which becomes the output value for the neuron and it passes to other neurons in the network. This is called multilayer perceptron (MLP). The number of parameters in a MLP with one hidden layer with  $h$  neurons and  $k$  inputs is given as;

$$h(k + 1) + h + 1 = h(k + 2) + 1 \quad (3.21)$$

By adjusting the weights on the connections between layers, the perceptron output can be “trained” to match a desired output. Weights ( $W_{ij}$ ) are determined by adding an error correction value to the old weight. The amount of correction is determined by multiplying the difference between the actual output ( $x[j]$ ) and target ( $t[j]$ ) values by a learning rate constant  $C$ . If the input node output ( $a[j]$ ) is a 1, that connection weight is adjusted, and if it sends 0, it has no bearing on the output and subsequently, there is no need for adjustment. The process can be represented as:

$$W_{ij(new)} = W_{ij(old)} + C(t_j - X_j)a_i \quad (3.22)$$

Where,  $C$  = learning rate. The training procedure is repeated until the network performance no longer improves. For the analyses done in research, a MLP neural network was employed, which is a feed-forward neural network using resilient propagation utilizing sigmoid activation functions. The number of iterations that the software runs has been configured to 50. Also, a large number of hidden nodes may increase training performance, but at the expense of

generalization and computation cost. Here, the performance was experimented with a number of hidden nodes and the nodes in a layer were chosen. The initial weights selected by the software are random and the final weights are the best weights obtained by error reduction at a convergence tolerance of 0.0001. The learning rate was set at 0.001 and the weight decay at 10. The activation function was a double sigmoid function as shown below:

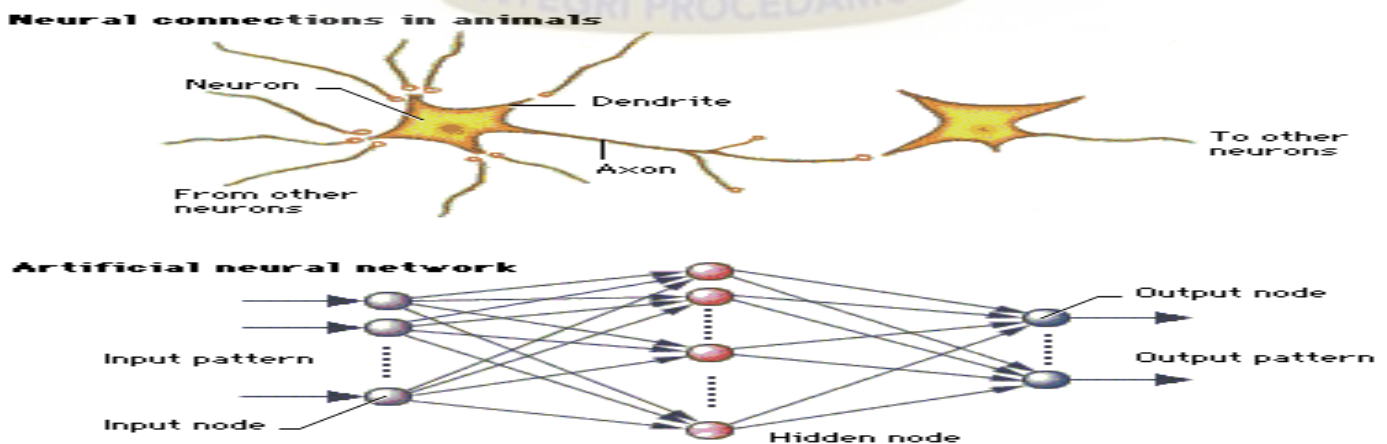
$$F(\text{sum}_j) = \frac{w_1}{1 + \exp(\text{sum}_j)} + \frac{w_2}{1 + \exp(-\text{sum}_j)} \quad (3.23)$$

where,  $\text{sum}_j$  is the scalar product of an input vector and weights to the node  $j$  either at a hidden layer or at the output layer and  $w_1$  and  $w_2$  are the initial weights.

**Predictor variables.** Predictors can be specified as factors (categorical) or covariates (scale).

**Categorical variable coding.** The procedure temporarily recoded the categorical predictors and dependent variables using one-of- $c$  coding for the duration of the procedure. If there are  $c$  categories of a variable, then the variable is stored as  $c$  vectors, with the first category denoted  $(1,0,\dots,0)$ , the next category  $(0,1,0,\dots,0)$ , ..., and the final category  $(0,0,\dots,0,1)$ .

Figure 3. 3: Feed Forward Architecture



The structure in Figure 3.3 is known as a feed forward architecture because the connections in the network flow forward from the input layer to the output layer without any feedback loops. The input layer contains the predictors which are Type of place of residence, Highest educational level of mother, Source of drinking water, type of toilet facility, Wealth index, Smokes cigarettes, Covered by health insurance, Marital status of mother, Sex of child, Prenatal Care, Whether child had diarrhea recently and Mothers age at time of birth and weight of baby

The hidden layer contains unobservable nodes, or units. The value of each hidden unit is some function of the predictors; the exact form of the function depends in part upon the network type and in part upon user-controllable specifications.

The output layer contains the response variable which was whether the child was alive or not. The MLP network allows a second hidden layer; in that case, each unit of the second hidden layer is a function of the units in the first hidden layer, and each response is a function of the units in the second hidden layer.

Partition Dataset was used to specify the method of partitioning the active dataset into training, testing, and holdout samples.

The training sample comprised the data records used to train the neural network; some percentage of cases in the dataset was assigned to the training sample in order to obtain a model.

The testing sample is an independent set of data records used to track errors during training in order to prevent overtraining. It is highly recommended that a training sample is created, and network training will generally be most efficient if the testing sample is smaller than the training sample.

The holdout sample is another independent set of data records which was used to assess the final neural network; the error for the holdout sample gives an "honest" estimate of the predictive ability of the model because the holdout cases were not used to build the model.

Cases based on relativity were randomly assigned to specify the relative number (ratio) of cases randomly assigned to each sample (training, testing, and holdout). The % column reported the percentage of cases that were assigned to each sample based on the relative numbers specified.

For example, specifying 7, 3, 0 as the relative numbers for training, testing, and holdout samples corresponds to 70%, 30%, and 0%. Specifying 2, 1, 1 as the relative numbers corresponds to 50%, 25%, and 25%; 1, 1, 1 corresponds to dividing the dataset into equal thirds among training, testing, and holdout.

Using partitioning variable to assigncases , a numeric variable was assigned to each case in the active dataset to the training, testing, or holdout sample. Cases with a positive value on the variable were assigned to the training sample, cases with a value of 0, to the testing sample, and cases with a negative value, to the holdout sample. Cases with system-missing values were excluded from the analysis.

### 3.4.1 Architecture

Architecture was the procedure used to specify the structure of the network. There are two types , automatic and custom architecture. Automatic architecture builds a network with one hidden layer. It was used to specify the minimum and maximum number of units allowed in the hidden layer, and was used to compute the "best" number of units in the hidden layer. Automatic architecture uses the default activation functions for the hidden and output layers. Custom architecture

selection gives control over the hidden and output layers and can be most useful when we know in advance what architecture we want or when we need to tweak the results of the Automatic architecture.

### **3.4.2 Hidden Layers**

The hidden layer contains unobservable network nodes (units). Each hidden unit is a function of the weighted sum of the inputs (explanatory variables). The function is the activation function, and the values of the weights were determined by the estimation algorithm. If the network contains a second hidden layer, each hidden unit in the second layer is a function of the weighted sum of the units in the first hidden layer. The same activation function is used in both layers.

### **3.4.3 Output Layer and Output**

The output layer contains the target (dependent) variables. The output displays information about the neural network, including the dependent variables (whether child is alive or not), number of input and output units, and number of hidden layers and units. The network diagrams were also displayed for interpretation.

Synaptic weights displayed the coefficient estimates that showed the relationship between the units in a given layer to the units in the following layer. The synaptic weights were based on the training sample even if the active dataset is partitioned into training, testing, and holdout data.

### **3.4.4 Network Performance**

The network displays results used to determine whether the model is "good" or "not". Charts in this group were based on the combined training and testing samples or only on the training

sample if there is no testing sample.

Model summary displayed a summary of the neural network results by partition and overall, including the error, the relative error or percentage of incorrect predictions, the stopping rule used to stop training, and the training time. The error is the sum-of-squares error when the identity, sigmoid, or hyperbolic tangent is applied to the output layer. Relative errors or percentages of incorrect predictions were displayed depending on the dependent variable measurement levels. Since our dependent variable (Child is alive or not) is categorical, then the average percentage of incorrect predictions is displayed. Relative errors or percentages of incorrect predictions were also displayed for the dependent variable.

#### **3.4.5 Classification of results**

A classification table for each categorical dependent variable by partition and overall was displayed. Each table gives the number of cases classified correctly and incorrectly for each dependent variable category. The percentage of the total cases that were correctly classified was also reported.

An ROC (Receiver Operating Characteristic) curve for each categorical dependent variable was also displayed. It also displays a table giving the area under each curve. For a given dependent variable, the ROC chart displayed one curve for each category. Since the dependent variable has two categories, then each curve treats the category at issue as the positive state versus the other category.

#### **3.4.6 Independent variable importance analysis**

It performs a sensitivity analysis, which computes the importance of each predictor in determining the neural network. The analysis is based on the combined training and testing

samples or only on the training sample if there is no testing sample. A table and a chart were created displaying the importance and normalized importance for each predictor

### 3.4.7 Probabilities and Pseudo-Probabilities

The Categorical dependent variable (whether child lives or dies) with cross-entropy error will have a predicted value for each category, where each predicted value is the probability that the case belongs to the category. The sum-of-squares error will have a predicted value for each category, but the predicted values cannot be interpreted as probabilities. The procedure saves these predicted pseudo-probabilities even if any are less than 0 or greater than 1, or the sum for a given dependent variable is not 1.

The ROC was created based on pseudo-probabilities. In the event that any of the pseudo-probabilities are less than 0 or greater than 1, or the sum for a given variable is not 1, they are first rescaled to be between 0 and 1 and to sum to 1. Pseudo-probabilities are rescaled by dividing by their sum. For example, if a case has predicted pseudo-probabilities of 0.50, 0.60, and 0.40 for a three-category dependent variable, then each pseudo-probability is divided by the sum 1.50 to get 0.33, 0.40, and 0.27. If any of the pseudo-probabilities are negative, then the absolute value of the lowest is added to all pseudo-probabilities before the above rescaling. For example, if the pseudo-probabilities are -0.30, 0.50, and 1.30, then first add 0.30 to each value to get 0.00, 0.80, and 1.60. Next each new value is divided by the sum 2.40 to get 0.00, 0.33, and 0.67.

### 3.5 The K-Nearest-Neighbor Method

Discriminant Analysis (DA), a multivariate statistical technique is commonly used to build a predictive/descriptive model of group discrimination based on observed predictor variables and to classify each observation into one of the groups. In DA multiple quantitative attributes are used to discriminate single classification variable. DA is different from the cluster analysis because prior knowledge of the classes, usually in the form of a sample from each class is required. The common objectives of DA are to investigate differences between groups, to discriminate groups effectively, to identify important discriminating variables and to perform hypothesis testing on the differences between the expected groupings and to classify new observations into pre-existing groups.

Non-parametric discriminant methods are based on non-parametric group-specific probability densities. Either a kernel or the k-nearest-neighbor method can be used to generate a non-parametric density estimate in each group and to produce a classification criterion. The performance of a discriminant criterion could be evaluated by non-parametric DA estimation, with the estimated group-specific densities and their associated prior probabilities, the posterior probability estimates of group membership for each class can be evaluated.

The k-nearest-neighbor method was the data mining technique employed in this study.

When the *k*-nearest-neighbor method is used, the Mahalanobis distances are estimated based on the pooled covariance matrix

Nearest Neighbor Analysis is a method for classifying cases based on their similarity to other cases. In machine learning, it was developed as a way to recognize patterns of data without requiring an exact match to any stored patterns, or cases. Similar cases are near each other and dissimilar cases are distant from each other. Thus, the distance between two cases is a measure of their dissimilarity. Cases that are near each other are said to be “neighbors.” When a new case

(holdout) is presented, its distance from each of the cases in the model is computed. The classifications of the most similar cases – the nearest neighbors – are tallied and the new case is placed into the category that contains the greatest number of nearest neighbors. One can specify the number of nearest neighbors to examine; this value is called  $k$ . The pictures (Figure 3.4) below show how a new case would be classified using two different values of  $k$ . When  $k = 5$ , the new case is placed in category 1 because a majority of the nearest neighbors belong to category 1. However, when  $k = 9$ , the new case is placed in category 0 because a majority of the nearest neighbors belong to category 0.

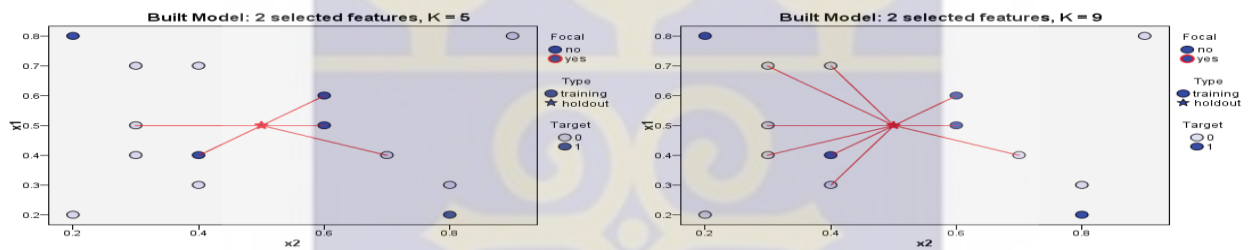


Figure 3. 4:  $K$ th nearest Neighbor

### 3.5.1 Classification

To demonstrate a  $k$ -nearest neighbor analysis, the task of classifying a new object (query point) among a number of known examples should be considered. This is shown in the figure below, which depicts the examples (instances) with the plus and minus signs and the query point with a red circle. The task is to estimate (classify) the outcome of the query point based on a selected number of its nearest neighbors. In other words, one wants to know whether the query point can be classified as a plus or a minus sign.

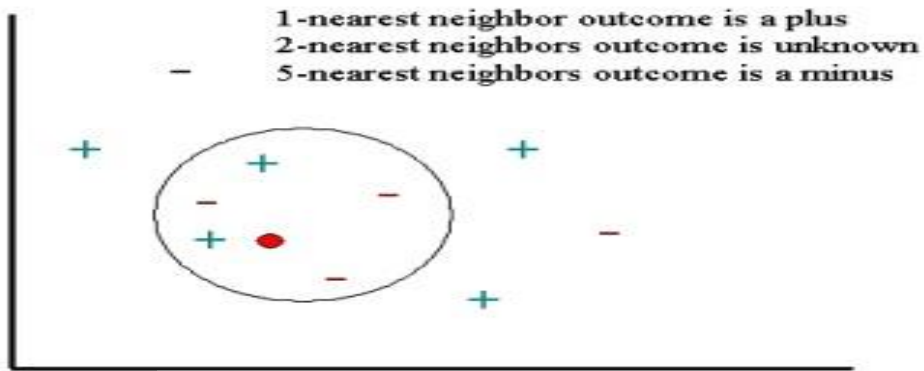
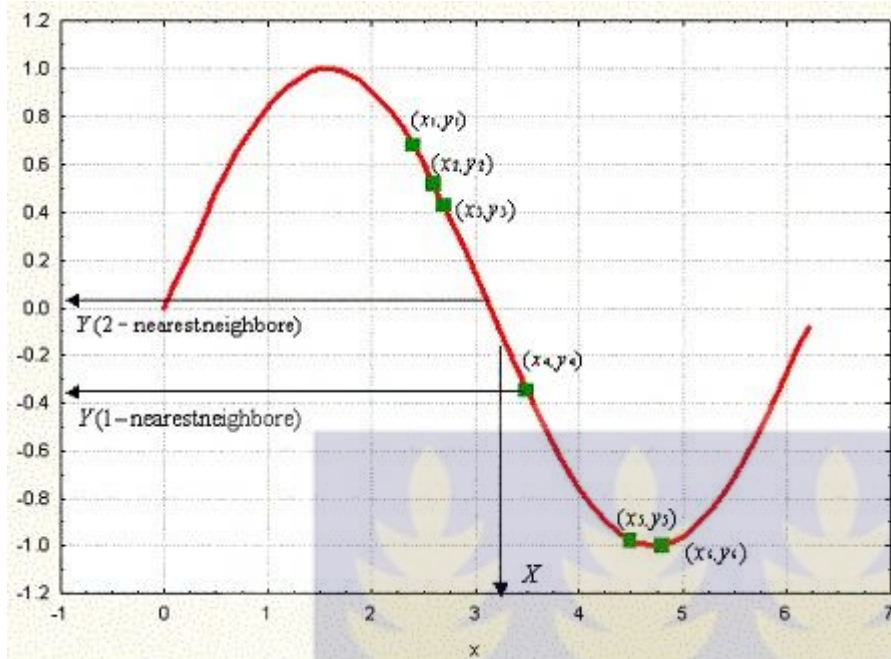


Figure 3.5 :Kth Nearest Neighbor

To proceed, consider the outcome of KNN based on 1-nearest neighbor. It is clear that in this case KNN will predict the outcome of the query point with a plus (since the closest point carries a plus sign). Now increase the number of nearest neighbors to 2, i.e., 2-nearest neighbors. This time KNN will not be able to classify the outcome of the query point since the second closest point is a minus, and so both the plus and the minus signs achieve the same score (i.e., win the same number of votes). For the next step, an increment in the number of nearest neighbors to 5 (5-nearest neighbors) should be made. This will define a nearest neighbor region, which is indicated by the circle shown in the figure above. Since there are 2 and 3 plus and minus signs, respectively, in this circle KNN will assign a minus sign to the outcome of the query point.



### 3.5.2 Regression

In this section the researcher will generalize the concept of  $k$ -nearest neighbors to include regression problems. Regression problems are concerned with predicting the outcome of a dependent variable given a set of independent variables. To start with, a consideration of the schematic shown above is made, where a set of points (green squares) are drawn from the relationship between the independent variable  $x$  and the dependent variable  $y$  (red curve). Given the set of green objects (known as examples) use the  $k$ -nearest neighbors method to predict the outcome of  $X$  (also known as query point) given the example set (green squares).

To begin with, consider the 1-nearest neighbor method as an example. In this case search the example set (green squares) and locate the one closest to the query point  $X$ . For this particular case, this happens to be  $x_4$ . The outcome of  $x_4$  (i.e.,  $y_4$ ) is thus then taken to be the answer for the outcome of  $X$  (i.e.,  $Y$ ). Thus for 1-nearest neighbor can be written as:

$$Y = y_4 \quad (3.24)$$

For the next step, consider the 2-nearest neighbor method. In this case, locate the first two closest points to  $X$ , which happen to be  $y_3$  and  $y_4$ . Taking the average of their outcome, the solution for  $Y$  is then given by:

$$Y = \frac{y_3 + y_4}{2} \quad (3.25)$$

The above discussion can be extended to an arbitrary number of nearest neighbors  $K$ . To summarize, in a  $k$ -nearest neighbor method, the outcome  $Y$  of the query point  $X$  is taken to be the average of the outcomes of its  $k$ -nearest neighbors.

### 3.5.3 Technical Details

STATISTICA k-Nearest Neighbors (KNN) is a memory-based model defined by a set of objects for which the outcome is known. The independent and dependent variables can be either continuous or categorical. For continuous dependent variables, the task is regression; otherwise it is a classification. Thus, *STATISTICA KNN* can handle both regression and classification tasks. Given a new case of dependent values (query point), one would like to estimate the outcome based on the *KNN* examples. *STATISTICA KNN* achieves this by finding  $k$  examples that are closest in distance to the query point, hence, the name *k-Nearest Neighbors*. For regression problems, *KNN* predictions are based on averaging the outcomes of the  $k$  nearest neighbors; for classification problems, a majority of voting is used. The choice of  $k$  is essential in building the *KNN* model. In fact,  $k$  can be regarded as one of the most important factors of the model that can strongly influence the quality of predictions. One appropriate way to look at the

number of nearest neighbors  $k$  is to think of it as a smoothing parameter. For any given problem, a small value of  $k$  will lead to a large variance in predictions. Alternatively, setting  $k$  to a large value may lead to a large model bias. Thus,  $k$  should be set to a value large enough to minimize the probability of misclassification and small enough (with respect to the number of cases in the example sample) so that the  $k$  nearest points are close enough to the query point. Thus, like any smoothing parameter, there is an optimal value for  $k$  that achieves the right trade off between the bias and the variance of the model. *STATISTICA KNN* can provide an estimate of  $k$  using an algorithm known as cross-validation (Bishop, 1995).

### 3.5.4 Cross-Validation

Cross-validation is a well-established technique that can be used to obtain estimates of model parameters that are unknown. Here a discussion of the applicability of this technique to estimating  $k$  is made. The general idea of this method is to divide the data sample into a number of  $v$  folds (randomly drawn, disjointed sub-samples or segments). For a fixed value of  $k$ , we apply the *KNN* model to make predictions on the  $v^{\text{th}}$  segment (i.e., use the  $v-1$  segments as the examples) and evaluate the error. The most common choice for this error for regression is sum-of-squared and for classification it is most conveniently defined as the accuracy (the percentage of correctly classified cases). This process is then successively applied to all possible choices of  $v$ . At the end of the  $v$  folds (cycles), the computed errors are averaged to yield a measure of the stability of the model (how well the model predicts query points). The above steps are then repeated for various  $k$  and the value achieving the lowest error (or the highest classification accuracy) is then selected as the optimal value for  $k$  (optimal in a cross-validation sense). The cross-validation is computationally expensive and should be prepared to let the algorithm run for

some time especially when the size of the examples sample is large. Alternatively,  $k$  can be specified. This may be a reasonable course of action should we have an idea of which value  $k$  may take (i.e., from previous *KNN* analyses that may have been conducted on similar data).

### 3.5.5 Distance Metric

As mentioned before, given a query point, *KNN* makes predictions based on the outcome of the  $K$  neighbors closest to that point. Therefore, to make predictions with *KNN*, one needs to define a metric for measuring the distance between the query point and cases from the examples sample. One of the most popular choices to measure this distance is known as Euclidean. Other measures include Euclidean squared, City-block, and Chebyshev:

$$D(x, p) = \sqrt{(x - p)^2} \quad \text{Euclidean}; \quad (x - p)^2 \quad \text{Euclidean squared} \quad (3.26) \quad D(x, p)$$

$$= \text{Abs}(x - p) \quad \text{Cityblock}; \quad \text{Max}(|x - p|) \quad \text{Chebyshev} \quad (3.27)$$

Where  $x$  and  $p$  are the query point.

### 3.5.6 Euclidean metric

Euclidean distance is the distance between two points in Euclidean space. Euclidean space was originally devised by the Greek mathematician Euclid around 300 B.C.E. to study the relationships between angles and distances. Euclidean geometry specifically applies to spaces of two and three dimensions. However, it can easily be generalized to higher order dimensions.

### 3.5.7 City block metric

The City block distance function computes the distance that would be traveled to get from one data point to the other if a grid-like path is followed. The Manhattan distance between two items is the sum of the differences of their corresponding components.

### 3.5.8 $k$ -Nearest Neighbor Predictions

After selecting the value of  $k$ , you can then make predictions. For regression,  $KNN$  predictions is the average of the  $k$ -nearest neighbors outcome.

$$y = \frac{1}{k} \sum_{i=1}^k y_i \quad (3.28)$$

where  $y_i$  is the  $i$ th case of the examples sample and  $y$  is the prediction (outcome) of the query point. In contrast to regression, in classification problems,  $KNN$  predictions are based on a voting scheme in which the winner is used to label the query. The  $k$  neighbors are allowed to have equal influence on predictions irrespective of their relative distance from the query point. An alternative approach (Shepard 1968) is to use arbitrarily large values of  $k$  (if not the entire prototype sample) with more importance given to cases closest to the query point. This is achieved using so-called distance weighting.

### 3.5.9 Distance Weighting

Since  $KNN$  predictions are based on the intuitive assumption that objects close in distance are potentially similar, it makes good sense to discriminate between the  $k$  nearest neighbors when making predictions, i.e., we let the closest points among the  $k$  nearest neighbors have a prominent role in affecting the outcome of the query point. This can be achieved by introducing

a set of weights  $W$ , one for each nearest neighbor, defined by the relative closeness of each neighbor with respect to the query point. Thus:

$$W(x, p_i) = \frac{\exp(-D(x, p_i))}{\sum_{i=1}^k \exp(-D(x, p_i))} \quad (3.29)$$

Where  $D(x, p_i)$  is the distance between the query point  $x$  and the  $i$ th case  $p_i$  of the example sample. It is clear that the weights defined in this manner above will satisfy:

$$\sum_{i=1}^k W(X_0, X_1) = 1 \quad (3.30)$$

Thus, for regression problems, we have

$$y = \sum_{i=1}^k W(x_0, x_1) y_i \quad (3.31)$$

For classification problems, the maximum of the above equation (3.31) is taken for each class variables. It is clear from the above discussion that when  $k > 1$ , one can naturally define the standard deviation for predictions in regression tasks using:

$$error\ bar = \pm \sqrt{\frac{1}{k-1} \sum_{i=1}^k (y - y_i)^2} \quad (3.32)$$

### 3.5.10 Neighbors

A full representation of the data set would involve 10 feature dimensions and one target group. The feature dimensions are the independent variables and the target group the dependent variable (whether child is alive).

### 3.5.11 Number of Nearest Neighbors ( $k$ )

The number of nearest neighbors was specified. A range of values was specified and the software was allowed to automatically choose the "best" number of neighbors within that range. Each value of  $k$  in the requested range was tested, and the  $k$ , and accompanying feature set, with the lowest error rate (or the lowest sum-of-squares error if the target is scale) was selected. The Euclidean metric and City block metric (Manhattan distance) were also computed.

Mahalanobis transformation which eliminates the correlation between the variables and standardizes the variance of each variable was also computed. When the Mahalanobis transformation is used in conjunction with Euclidean distance, it is called Mahalanobis distance.

### 3.5.12 Feature Selection

Feature selection is based on the wrapper approach of Cunningham and Delany (2007) and uses forward selection which starts from  $J_{\text{forced}}$  features which are entered into the model. Further features are chosen sequentially; the chosen feature at each step is the one that causes the largest decrease in the error rate or sum-of squares error.

Let  $S_j$  represent the set of  $J$  features that are currently chosen to be included,  $S_j^c$  represents the

set of remaining features and  $e_j$  represents the error rate or sum-of-squares error associated with the model based on  $S_j$ .

### **3.5.13 Stopping Criterion**

At each step, the feature whose addition to the model results in the smallest error (computed as the error rate for a categorical target and sum of squares error for a scale target) is considered for inclusion in the model set. Forward selection continued until the specified condition was met.

### **3.5.14 Minimum change in absolute error ratio**

The algorithm stops when the change in the absolute error ratio indicates that the model cannot be further improved by adding more features. Decreasing values of the minimum change will tend to include more features, at the risk of including features that don't add much value to the model. Increasing the value of the minimum change will tend to exclude more features, at the risk of losing features that are important to the model.

### **3.5.15 Partitions**

The dataset was divided into training and holdout sets. It specifies the method of partitioning the active dataset into training and holdout samples.

The training sample comprised the data records used to train the nearest neighbor model; some percentages of cases in the dataset were assigned to the training sample in order to obtain a model. The holdout sample is an independent set of data records used to assess the final model; the error for the holdout sample gives an "honest" estimate of the predictive ability of the model because the holdout cases were not used to build the model.

A numeric variable was assigned to each case in the active dataset to the training or holdout

sample. Cases with a positive value on the variable were assigned to the training sample, cases with a value of 0 or a negative value, to the holdout sample. Cases with a system-missing value were excluded from the analysis.

### 3.5.16 Cross-Validation Folds

V-fold cross-validation was used to determine the "best" number of neighbors. Cross-validation divides the sample into a number of subsamples, or folds. Nearest neighbor models are then generated, excluding the data from each subsample in turn. The first model is based on all of the cases except those in the first sample fold, the second model is based on all of the cases except those in the second sample fold, and so on. For each model, the error was estimated by applying the model to the subsample excluded in generating it. The "best" number of nearest neighbors is the one which produces the lowest error across folds.

### 3.5.17 Output

The output displays a classification table for each categorical dependent variable by partition and overall. Each table gives the number of cases classified correctly and incorrectly for each dependent variable category. The percentage of the total cases that were correctly classified was also reported.

The model view included  $k$  nearest neighbors and distances for focal cases, classification of categorical response variables, and an error summary. Graphical output in the model view included a selection error log, feature importance chart, feature space chart, peers chart, and quadrant map.

### **The Feature Space Chart**

The feature space chart is an interactive graph of the feature space (or a subspace, if there are more than 3 features). Each axis represents a feature in the model, and the location of points in the chart show the values of these features for cases in the training and holdout partitions.

### **Variable Importance Chart**

Typically the researcher will want to focus the modeling efforts on the variables that matter most and consider dropping or ignoring those that matter least. The variable importance chart helps to achieve this by indicating the relative importance of each variable in estimating the model. Since the values are relative, the sum of the values for all variables on the display is 1.0. Variable importance does not relate to model accuracy. It just relates to the importance of each variable in making a prediction, not whether or not the prediction is accurate.

### **Peers Chart**

This chart displays the focal cases and their k nearest neighbors on each feature and on the target. It is available if a focal case is selected in the Feature Space.

### **Nearest Neighbor Distances**

This table displays the k nearest neighbors and distances for focal cases only. It is available if a focal case identifier is specified on the Variables tab, and only displays focal cases identified by this variable.

### **Quadrant Map**

This chart displays the focal cases and their k nearest neighbors on a scatterplot (or dotplot,

depending upon the measurement level of the target) with the target on the y-axis and a scale feature on the x-axis, paneled by features.

### **k Selection Error Log**

Points on the chart display the error (either the error rate or sum-of-squares error, depending upon the measurement level of the target) on the y-axis for the model with the number of nearest neighbors (k) on the x-axis.

### **3.5.18 Variable Importance**

Typically, the focus is aimed at modeling efforts on the variables that matter most and considering dropping or ignoring those that matter least. The variable importance chart helped by indicating the relative importance of each variable in estimating the model. Since the values are relative, the sum of the values for all variables on the display is 1.0.

### **3.5.19 Classification Table**

This table displays the cross-classification of observed versus predicted values of the target, by partition. The Holdout partition contains holdout cases with missing values on the target. These cases contribute to the Holdout Sample: Overall Percent values but not to the Percent Correct values.

### **3.5.20 Error Summary**

It displayed the error associated with the model; the error rate for a categorical target.

$$\text{Error rate} = \left(1 - \sum_{j=1}^J \frac{N_j}{N}\right) * 100 \quad (3.33)$$

Where  $N_j$  is the number of cases which belong to class  $j$  and are correctly classified

$asj.N$  is the Total number of cases

### 3.6 Comparison of Models

In order to evaluate the most appropriate technique, the predicted results for each model will be used. The predictive results can be determined by the correct classification rate (CCR) for the testing set.

There are two possible prediction errors in a two-class problem: false positives and false negatives and the performance of a binary classifier is normally summarized in a confusion or error matrix (Table ) that cross-tabulates observed and predicted positives/negatives as four numbers: a, b, c and d.

Columns are the actual groups, rows are the predictions. a = true positives, b = false positive, c = false negatives and d = true negative.  $n = a + b + c + d$ .

	Actual +	Actual -
Predicted +	a	b
Predicted -	c	d

Accuracy measures can then be calculated from the confusion matrix. Some of the measures that can be calculated are Overall Diagnostic Power, Prevalence test, Sensitivity test, Correct Classification Rate, Specificity test, Positive Predictive Power, Misclassification Rate, Positive Predictive Power, Positive Predictive Power and Kappa test.

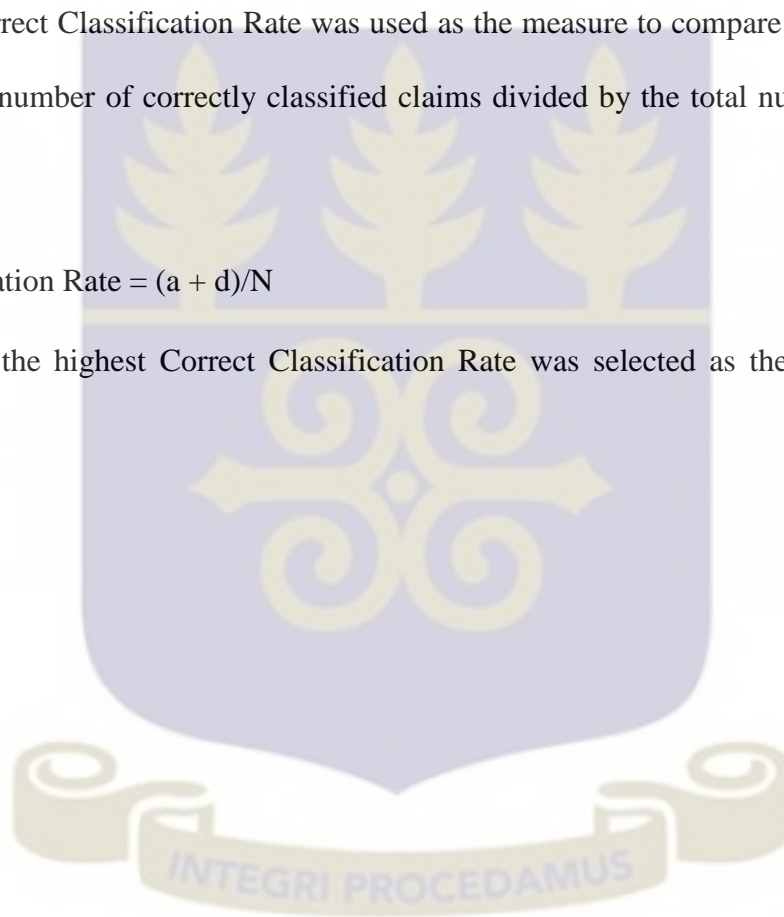
These measures have different characteristics. For example, sensitivity (proportion of correctly classified positive cases), takes no account of the false positives while specificity measures the false positive errors. Some of the measures are sensitive to the prevalence of positive cases and

some can be used to measure the improvement over chance. Kappa (K, proportion of specific agreement), is often used to assess improvement over chance, with  $K < 0.4$  indicating poor agreement. However, K is sensitive to the sample size and it is unreliable if one class dominates. Although the related NMI measure does not suffer from these problems it shows non-monotonic behaviour under conditions of excessive errors.

In this study, Correct Classification Rate was used as the measure to compare the three different models. It is the number of correctly classified claims divided by the total number of claims in the data set.

$$\text{Correct Classification Rate} = (a + d)/N$$

The model with the highest Correct Classification Rate was selected as the most appropriate model



## **CHAPTER FOUR**

### **RESULTS AND DISCUSSION**

#### **4.1 Introduction**

This chapter outlines the results obtained in the study and provides a discussion of these results. Three different modeling techniques were considered in this research. The results obtained by the analysis using logistic regression were compared with the results obtained by neural networks and nearest neighbor data mining techniques. The data mining software IBM SPSS Version 20 was used for building the models. The chapter is divided into five sections; preliminary analyses, logistic regression, neural networks, K- nearest neighbor and comparison of techniques.

#### **4.2 Preliminary Analyses**

A preliminary analysis was carried out to reveal certain characteristics of the data. Table 4.1 reveals the frequency and percentage distribution of the explanatory variables against whether a child is alive or dead. It describes the proportion of responses for each variable. For example majority of the respondents (67 %) hail from the rural areas of Ghana. Of the 67%, there was a 12% reported child deaths from the respondents of the survey. The other variables in the table are interpreted as such.

**Table 4.1: Frequency Distribution of Variables used in the study**

		Child is alive?				Total	
		Count	Subtable Valid N %	Count	Subtable Valid N %	Count	Subtable Valid N %
Type of place of residence	Urban	3253	27.4%	665	5.6%	3918	33.0%
	Rural	6547	55.1%	1423	12.0%	7970	67.0%
Highest educational level	No education	4192	35.3%	920	7.7%	5112	43.0%
	Primary	2126	17.9%	538	4.5%	2664	22.4%
	Secondary	3301	27.8%	606	5.1%	3907	32.9%
	Higher	172	1.4%	23	0.2%	195	1.6%
Source of drinking water	PIPED WATER	3132	26.3%	692	5.8%	3824	32.2%
	NON PIPED WATER	6667	56.1%	1396	11.7%	8063	67.8%
Type of toilet facility	FLUSH TOILET	717	6.0%	148	1.2%	865	7.3%
	PIT LATERINE	5435	45.8%	1300	11.0%	6735	56.7%
	NO FACILITY	3631	30.6%	638	5.4%	4269	36.0%
Covered by health insurance	No	5758	48.5%	1319	11.1%	7077	59.6%
	Yes	4026	33.9%	767	6.5%	4793	40.4%
Marital status	Never married	187	1.6%	24	0.2%	211	1.8%
	Currently married	8496	71.5%	1784	15.0%	10280	86.5%
	Formerly married	1117	9.4%	280	2.4%	1397	11.8%
Sex of child	Male	4871	41.0%	1124	9.5%	5995	50.4%
	Female	4929	41.5%	964	8.1%	5893	49.6%
Alcohol consumption	No	7299	61.4%	1630	13.7%	8929	75.1%
	Yes	2496	21.0%	458	3.9%	2954	24.9%

Wealth index	Rich	2869	24.1%	516	4.3%	3385	28.5%
	Middle	1805	15.2%	406	3.4%	2211	18.6%
	Poor	5126	43.1%	1166	9.8%	6292	52.9%

Source : GDHS,2008

Table 4.2: Pearson Chi-Square Tests

		Child is alive?
Type of place of residence	Chi-square	1.410
	df	1
	Sig.	.235
Highest educational level	Chi-square	29.251
	df	3
	Sig.	.000*
Source of drinking water	Chi-square	1.097
	df	1
	Sig.	.295
Type of toilet facility	Chi-square	34.382
	df	2
	Sig.	.000*
Covered by health insurance	Chi-square	13.701
	df	1
	Sig.	.000*
Marital status	Chi-square	11.825
	df	2
	Sig.	.003*
Sex of child	Chi-square	11.729
	df	1
	Sig.	.001*
Alcohol consumption	Chi-square	11.596

	df	1
	Sig.	.001*
Wealth index	Chi-square	17.628
	df	2
	Sig.	.000*

Results are based on nonempty rows and columns in each innermost subtable.

\*. The Chi-square statistic is significant at the .05 level.

Table 4.3: Age of mother at birth

Cases	Mean age	Median age	Minimum age	Maximum age
11888	36.62	37	15	45

Also Table 4.2 is a chi square test to ascertain whether there is an existence of association between the independent variables and the dependent variable. A P – value  $\leq 0.05$  is deemed to be significant and will support the hypothesis that there exists an association between the variables. Therefore from Table 4.2 all the variables with the exception of type of place of residence and source of drinking water were significant and hence have an association with the dependent variable(Is child alive).

Table 4.3 presents a summary statistics of the only quantitative variable in the study: the age of mother at birth of child. The average age was 37 years with the minimum and maximum age of 15 and 45 years respectively.

### 4.3 Logistic Regression Model

The outcome variable was whether a child is alive or dead. The outcome variable for the model was defined as follows;

$$y_i(\text{is child alive?}) = \{1 = \text{Child}^i \text{ dies ; } 0 = \text{Child}^i \text{ is alive } \}$$

Table 4.4: Case Processing Summary

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	11,836	99.6
	Missing Cases	52	.4
	Total	11,888	100.0
Unselected Cases		0	.0
Total		11,888	100.0

Table 4.5: Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	10768.032 <sup>a</sup>	.020	.034

Under Model Summary we see that the -2 Log Likelihood statistic is 10768.032. This statistic measures how poorly the model predicts the decisions, the smaller the statistic the better the model. The Model Summary table also indicates how much of the variance in the dependent variable is explained by the model. It can be observed from Table 4.5 that, between 2.0% and 3.4% of the variance in predicting mortality likelihood among under five children was explained by the significant independent variables. A more useful measure to assess the utility of a logistic regression model is classification accuracy, which compares predicted group membership based on the logistic model to the actual, known group membership, which is the value for the

dependent variable. To characterize our model as useful, we compare the overall percentage accuracy rate produced by the software at the last step in which variables are entered. The overall accuracy rate was 81.00% (see table 4.6)

Table 4.6: Classification Table

			Child is alive?		Percentage Correct
			Yes	No	
Step 1	Child is alive?	Yes	8187	1566	83.94
		No	683	1400	67.21
Overall Percentage					81.00

Table 4.7: Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	244.581	15	.000
	Block	244.581	15	.000
	Model	244.581	15	.000

Table 4.7 displays the results for the omnibus tests of model coefficients. The presence of a relationship between the dependent variable and combination of independent variables is based on the statistical significance of the model chi-square at step 1 after the independent variables have been added to the analysis. In this analysis, the probability of the model chi-square (244.581) was  $< 0.001$ , less than or equal to the level of significance of 0.05. The null hypothesis that there is no significant difference between the model with only a constant and the model with independent variables was rejected. The existence of a relationship between the independent variables and the dependent variable was supported.

Classification Table <sup>a,b</sup>					
Step 1	Observed		Predicted		Percentage Correct
	Child is alive?		Child is alive?		
			Yes	No	
	Yes		8187	1566	83.94
	No		683	1400	67.21
	Overall				81.00
	Percentage				

Table 4.8: Variables in the Equation of the Logistic Regression Model

Variables in the Equation						
	$\beta$	S.E.	Wald	df	P-value	Odds
PLACE_OF_RESIDENCE(1)	.050	.073	.475	1	.491	1.051
EDUCATIONAL_LEVEL			25.751	3	.000	
EDUCATIONAL_LEVEL(1)	.515	.007	4.726	1	.030	1.674
EDUCATIONAL_LEVEL(2)	.634	.009	7.183	1	.007	1.501
EDUCATIONAL_LEVEL(3)	.312	.024	1.797	1	.180	1.366
Source_of_drinking_water(1)	.129	.065	4.002	1	.045	1.138
Toilet_facility			57.315	2	.000	
Toilet_facility(1)	.617	.122	25.431	1	.000	1.854
Toilet_facility(2)	.457	.062	54.250	1	.000	1.579
Wealth_index			30.675	2	.000	
Wealth_index(1)	-.492	.089	30.321	1	.000	.612
Wealth_index(2)	-.244	.075	10.666	1	.001	.784
Age_of_mother	.031	.003	82.199	1	.000	1.031
health_insurance(1)	.136	.052	6.952	1	.008	1.146
Marital_status			.949	2	.622	
Marital_status(1)	-.190	.234	.659	1	.417	.827
Marital_status(2)	-.056	.074	.565	1	.452	.946
Sex_of_child(1)	.180	.049	13.471	1	.000	1.197
alcohol_consumption(1)	.195	.060	10.498	1	.001	1.215
Constant	-3.625	.292	153.985	1	.000	.027

Table 4.8 presents the results in the final logistic regression model. It lists the estimated model parameters, their standard errors, odds ratio, Wald tests, and associated *P*-values. The parameter estimates are the estimated coefficients of the fitted logistic regression model.

The coefficient of any variable is deemed to be significant if  $P \text{ value} \leq 0.05$ . Hence the significant values in the model are Mothers educational level, Age of mother at birth, Type of toilet facility used, Wealth index of family, whether mother is registered with health insurance, Sex of child and alcohol consumption.

For Odds Ratio any, value greater than one in the table increases the likelihood that a child will die. An Odds Ratio value of less than one decreases the likelihood that a child will die.

The probability of the Wald statistic for the independent variable educational level of mother was less than or equal to the level of significance of 0.05. The null hypothesis that the  $\beta$  coefficient for educational level of mother was equal to zero was rejected. This means that the educational level of mother was significant in predicting mortality in children. The value of the Odds for no education (relative to higher education) is 1.674, which implies that a unit increase in no education (higher education as reference variable) will increase the odds that the respondent's child in the survey will die by 67.4% holding all other variables constant. Also a unit increase in primary education (relative to higher education) will increase the odds that a child will die by 50.1%. Secondary education (relative to higher education) was found to be insignificant in the model.

The probability of the Wald statistic for the independent variable toilet facility used (0.00) was less than or equal to the level of significance of 0.05. The null hypothesis that the  $\beta$  coefficient was equal to zero for toilet facility used was rejected. The value of odds for flush toilet (no toilet

facility as reference variable) is 85.4%.Pit Latrine(no toilet facility as reference variable) will increase the odds that a child will die by 57.9%

The other significant variables in table are also interpreted as such.

Multicollinearity in the logistic regression solution is detected by examining the standard errors for the  $\beta$  coefficients. A standard error larger than 2.0 indicates numerical problems, such as multicollinearity among the independent variables, From Table 4.8 we can deduce that none of the independent variables had a standard error greater than 2.0 hence there is no evidence of multicollinearity. The correlation coefficient matrix(see Appendix) also suggests that multicollinearity was not an issue in this model.

Table 4.9: Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	3.229	8	.919

From Table 4.9 , since the p – value, 0.919, is greater than the significance level,  $\alpha = 0.05$ , we fail to reject the null hypothesis ( $H_0$ ) and conclude that there is enough evidence to show that the hypothesized model fits the data set used in predicting mortality among under five children in Ghana.

#### 4.4 Neural Network

Table 4.10: Case Processing Summary

		N	Percent
Sample	Training	8264	69.8%
	Testing	2428	20.5%
	Holdout	1144	9.7%
Valid		11836	100.0%
Excluded		52	
Total		11888	

The dataset was partitioned into training, testing, and holdout samples. The training sample comprises the data records used to train the neural network; 69.8% of cases in the dataset were assigned to the training sample in order to obtain a model.

The testing sample is an independent set of data records used to track errors during training in order to prevent overtraining. It is highly recommended that we create a training sample, and network training will generally be most efficient if the testing sample is smaller than the training sample. A percentage of 20.5 of cases in the dataset was assigned to the testing sample.

The holdout sample is an independent set of data records used to assess the final model; the error for the holdout sample gives an "honest" estimate of the predictive ability of the model because the holdout cases were not used to build the model. 9.7% of cases was assigned to the holdout sample.

Figure 4.1: Feedforward architecture with one hidden layer

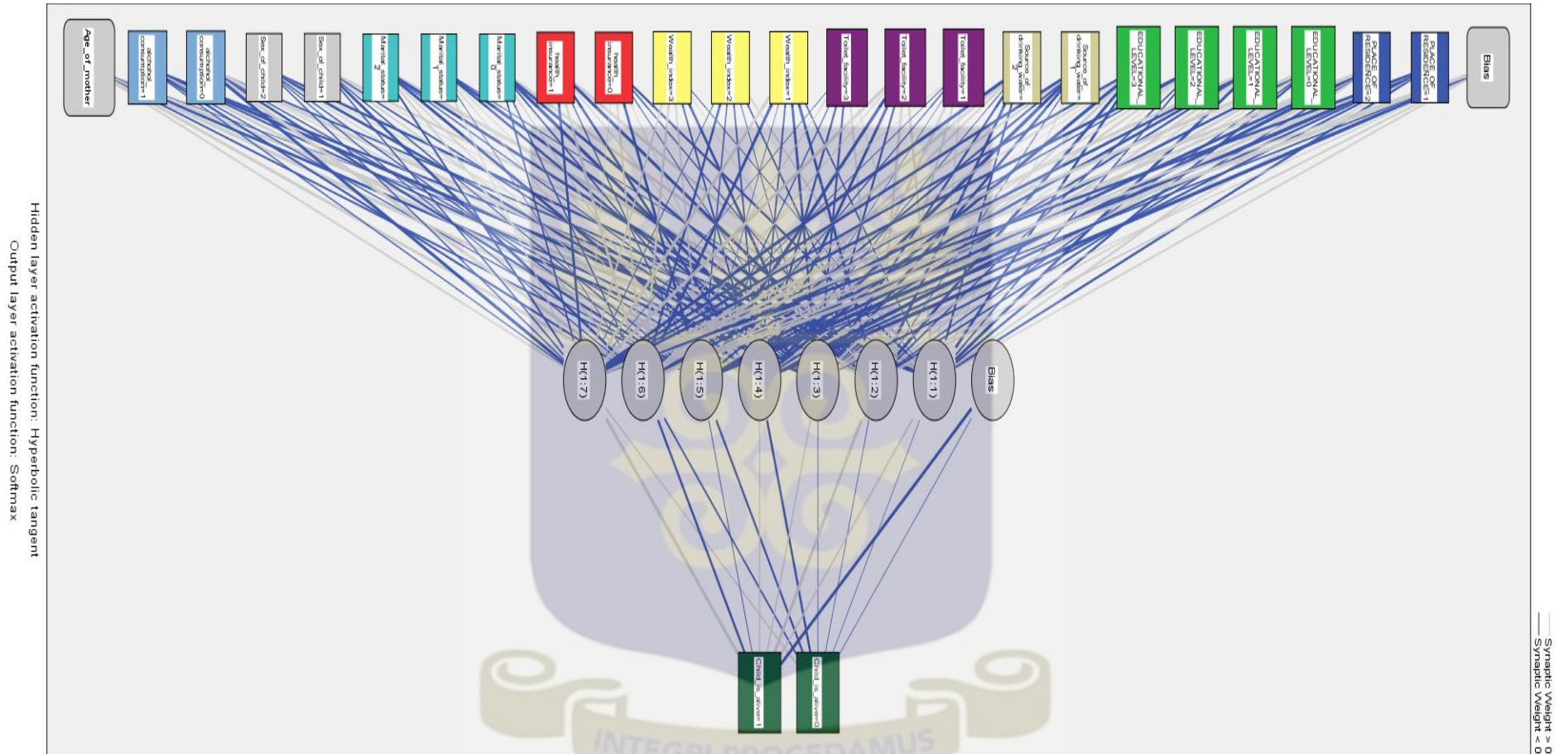


Table 4.11: Model Summary

Training	Cross Entropy Error	3831.846
	Percent Incorrect Predictions	20.6%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error <sup>a</sup>
	Training Time	0:00:00.58
Testing	Cross Entropy Error	1141.422
	Percent Incorrect Predictions	19.4%
Holdout	Percent Incorrect Predictions	20.5%

Table 4.12: Classification Table

Sample	Observed	Predicted		
		Yes	No	Percent Correct
Training	Yes	5615	1206	82.32%
	No	500	943	65.35%
	Overall Percent	74.00%	26.00%	79.36%
Testing	Yes	1670	313	84.22%
	No	155	290	65.17%
	Overall Percent	75.16%	24.84%	80.72%
Holdout	Yes	785	164	82.72%
	No	71	125	63.78%
	Overall Percent	74.76%	25.24%	79.48%

Dependent Variable: Child is Alive

The model summary shows the percentage of incorrect predictions is roughly equal across all the samples which is a positive sign.

The classification table shows that, using 0.5 as the pseudo-probability cutoff for classification, the network does fairlywell at predicting the likelihood of death or otherwise in children under

five years. Of the cases used to create the model, 5615 of the 6851 children alive were classified correctly. 943 of the 1443 child deaths were classified correctly. Overall, 79.4% of the training cases were classified correctly, corresponding to the 20.6% incorrect shown in the model summary table. The testing sample also reveals that 1670 of the 1983 children alive were classified correctly while 290 of the 445 child deaths were classified correctly. Overall, 80.7% of the testing cases were correctly classified by the model. The overall classification for the holdout sample was 79.5% which is very close to both the training and testing samples. This suggests that overall, the model was about 82.7% correct classification for children alive while that of children who are not alive was about 65.2%.

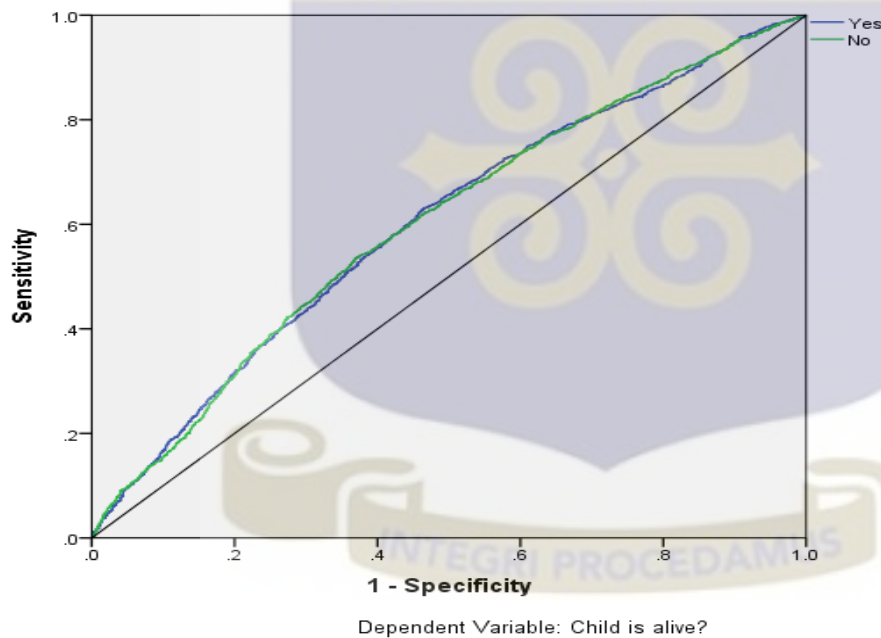


Figure 4.2: ROC Curve of the Neural Network Model

The ROC curve gives a visual display of the sensitivity and specificity for all possible cutoffs.

Table 4.13: Area under the Curve

		Area
Child is alive?	Yes	.701
	No	.701

From Table 4.12, which is a numerical summary of the ROC for a randomly selected child who is alive or a randomly selected child who is dead, there is a 0.701 probability that the model-predicted pseudo-probability will be higher for a child who is alive than for a child who is dead.

Table 4.14: Independent Variable Importance

	Importance	Normalized Importance
Age of mother	0.304	100.00%
Highest educational level	0.17	55.80%
Type of toilet facility	0.121	39.60%
Alcohol consumption	0.084	27.50%
Wealth index	0.073	24.00%
Source of drinking water	0.068	22.30%
Covered by health insurance	0.061	20.20%
Type of place of residence	0.058	19.20%
Sex of child	0.035	11.40%
Marital status	0.027	8.70%



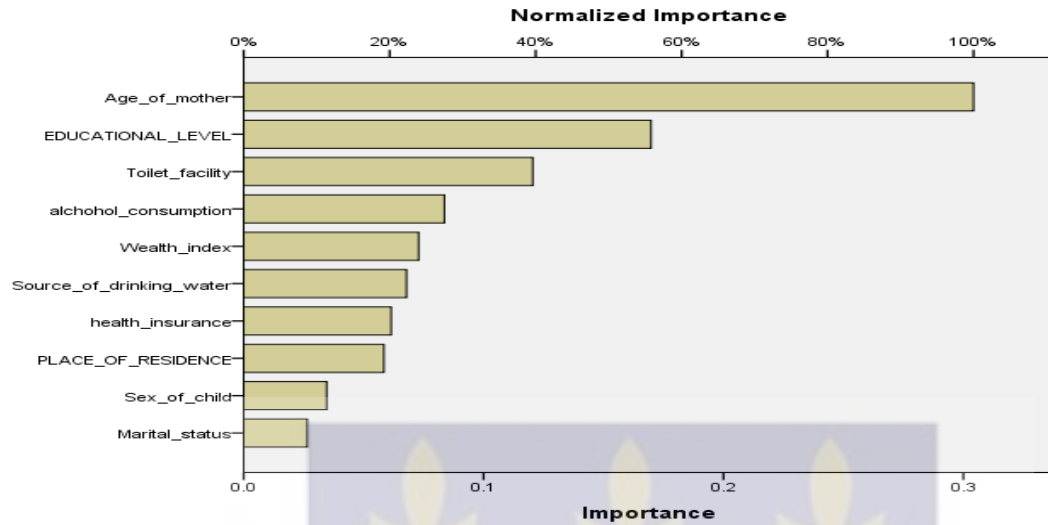


Figure 4.3: Normalized Importance Chart

The importance chart is simply a bar chart of the values in the importance table, sorted in descending value of importance. It appears that characteristics of the mother (age of mother, mothers' educational level and mothers alcohol consumption) and sanitation issues (type of toilet facility) have the greatest effect on how the network classifies whether a child live or die; what neural network cannot reveal the "direction" of the relationship between these variables.



## 4.5 Nearest Neighbor Analysis

Table 4.15: Case Processing Summary

Case Processing Summary			
		N	Percent
Sample	Training	8282	70.0%
	Holdout	3554	30.0%
Valid		11836	100.0%
Excluded		52	
Total		11888	

The dataset was divided into training and holdout sets. It specifies the method of partitioning the active dataset into training and holdout samples.

The training sample comprised the data records used to train the nearest neighbor model; 70% of cases in the dataset were assigned to the training sample in order to obtain a model.

The holdout sample is an independent set of data records used to assess the final model; the error for the holdout sample gives an "honest" estimate of the predictive ability of the model because the holdout cases were not used to build the model. 30% of cases were assigned to the holdout sample.

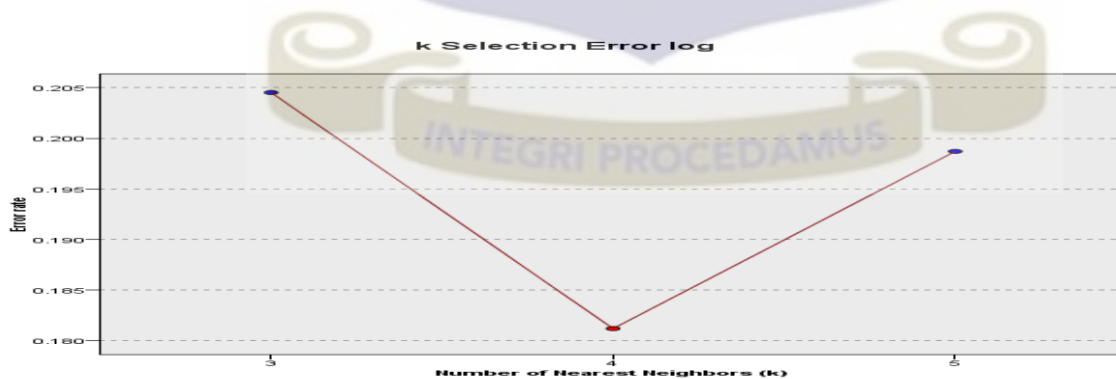


Figure 4.4: K Selection Error Log Chart

Table 4.16: Error Summary for Nearest Neighbor

Partition	Percent of Records Incorrectly Classified
Training	16.8%
Holdout	16.6%

The statistical software was allowed to automatically select the value of k. k = 4 was selected because from Figure 4.4, it had the lowest error rate.

The error summary in Table 4.15 shows the percentage of incorrect predictions is roughly equal across both the training and holdout samples which is a good sign that the model is a good fit.

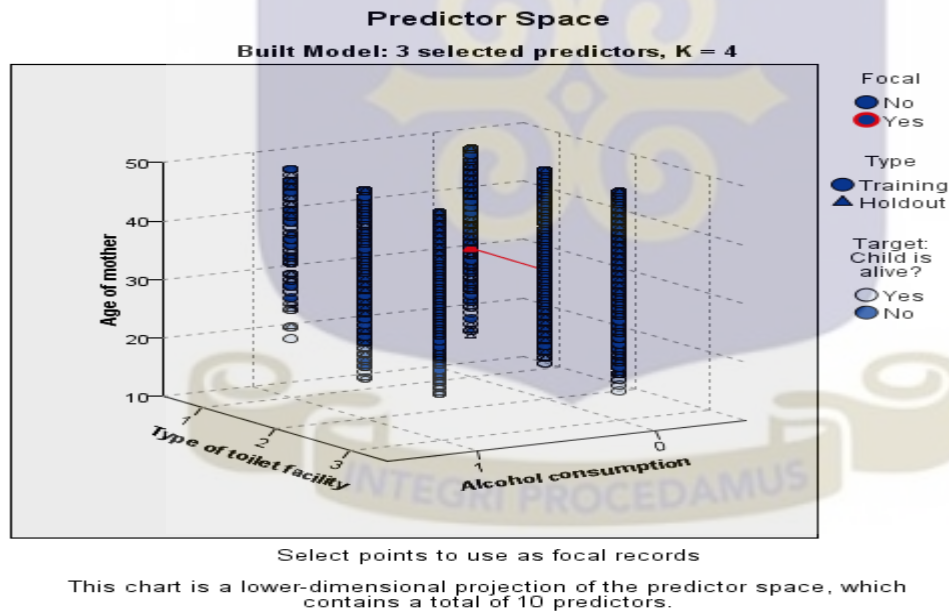


Figure 4.5: Predictor Space

In Nearest Neighbor analysis the predictor variables are referred to as feature variables and the dependent variable is referred to as the target variable. Figure 4.4 displays a feature or

predictor space chart with 10 dimensions based on the number of feature variables. In the chart similar cases are near each other and dissimilar cases are distant from each other. The Euclidean metric was used in the computation of distances within the chart.

A focal case (point) was randomly selected and marked red in the chart. The analysis will then be based on the focal point randomly selected. From any selected point, the four nearest neighbors to that point will be selected using the Euclidean metric and based on majority vote, nearest neighbors with similar characteristics will be selected as the characteristic of the selected focal point. We can use this concept to make predictions.

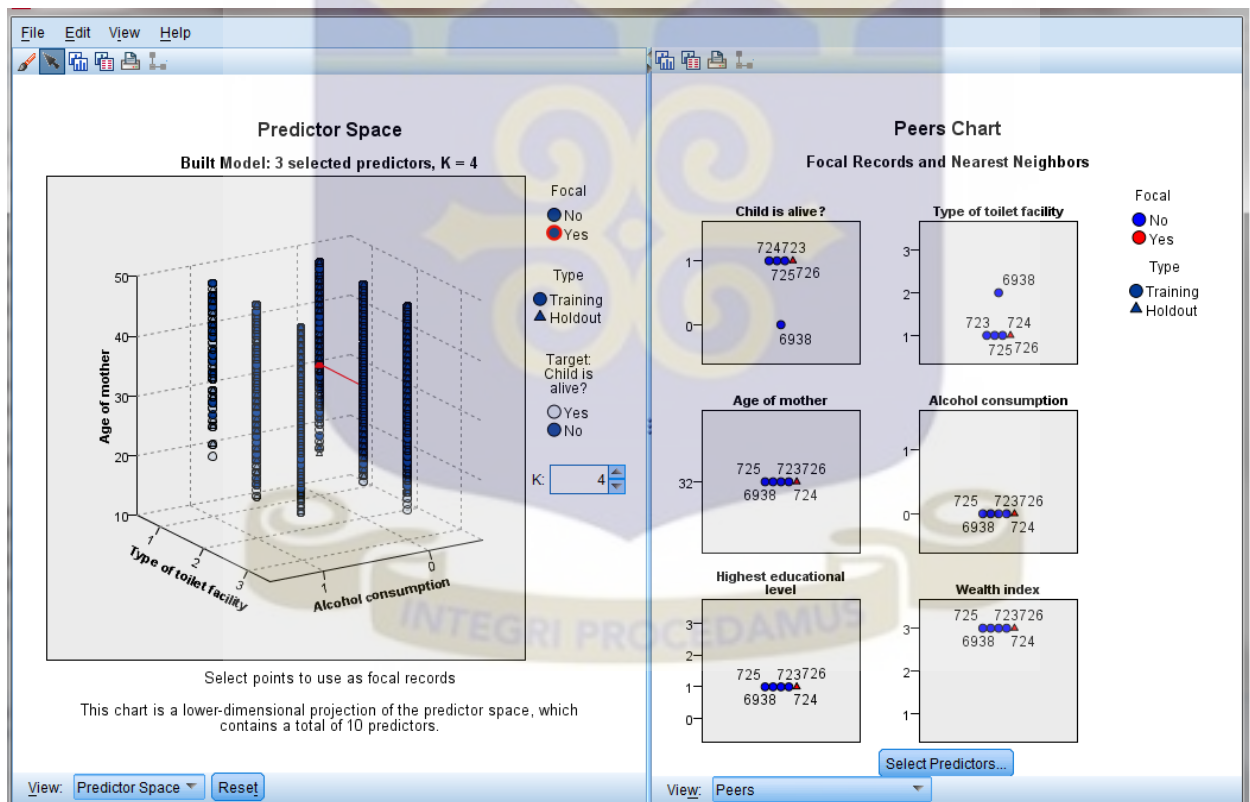


Figure 4.6: Peers Chart

Table 4.17:4 Nearest Neighbors and Distances

Focal Record	Displayed for Initial Focal Records							
	Nearest Neighbors				Nearest Distances			
	1	2	3	4	1	2	3	4
<b>726</b>	724	723	725	6938	0.436	0.436	0.436	0.454

Table 4.18: Summary of the Peers Chart

OBS.	Child	Age	Educ.	Toilet	Alcohol	Wealth
<b>726</b>	<b>Dead</b>	<b>32</b>	<b>primary</b>	<b>None</b>	<b>No</b>	<b>poor</b>
723	Dead	32	Primary	None	No	poor
724	Dead	32	Primary	None	No	poor
725	Dead	32	Primary	None	No	poor
6938	Alive	32	Primary	Pit Laterine	No	poor

Figure 4.6 is a peers chart. It displays how a randomly selected focal point and its four nearest neighbors are represented in the predictor space taking into consideration the feature variables (independent).

Table 4.17 then presents the observation (point) numbers of the focal point and its nearest neighbors. Their Euclidean distances from the focal point are also computed.

Table 4.18 summarizes how the data is presented in the peers chart. It can clearly be seen that the nearest neighbors of the randomly selected focal point have confirmed the characteristics of that focal point. This is because three (723, 724 and 725) out of the four nearest neighbors have the same attributes as the focal point. An alcohol free mother from a poor home whose age at birth of child under five was 32 years. Her highest level of education was primary school. She does not have a toilet facility in her house. The end

result was that she lost her child under five years. Points 723,724 and 725 have the same attributes. This means without knowing the attributes of the focal point, its nearest neighbors can be used to predict whether a child will live or die taking into consideration the attributes of the subject. This can help to predict what an outcome would be given certain characteristics of a subject. The Nearest Neighbors of any point selected in the feature space using the majority vote can provide information about the attribute of the selected point.

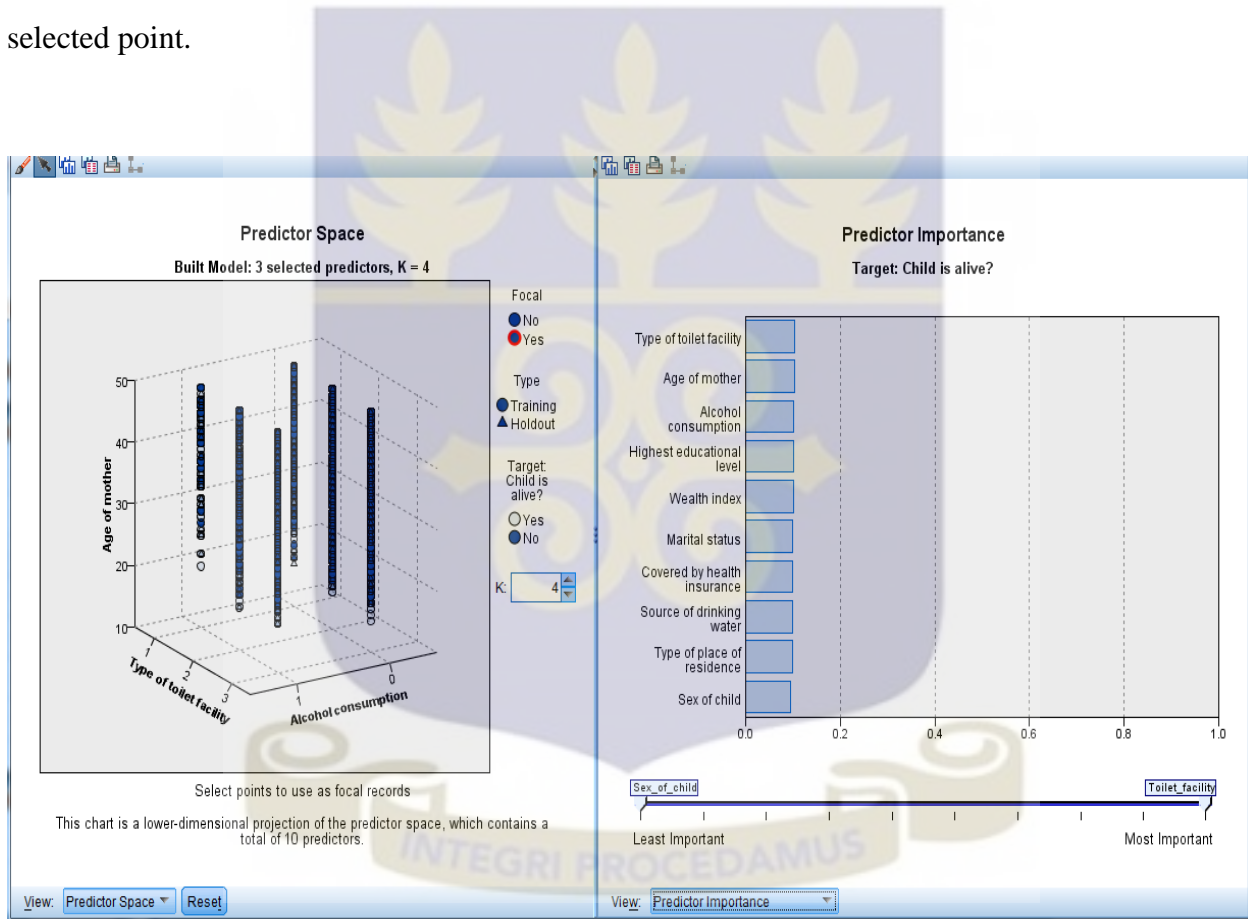


Figure 4.7: Predictor Importance Chart for Nearest Neighbor

From Figure 4.7 the characteristics of the mother (age of mother, educational level, alcohol consumption and Wealth index) played an important role in modeling mortality in children

under five .Sanitation(type of toilet facility) was the most important predictor of mortality.

Table 4.19: Classification Table for Nearest Neighbor

<b>Partition</b>	<b>Observed</b>	<b>Predicted</b>		
		<b>Yes</b>	<b>No</b>	<b>Percent Correct</b>
<b>Training</b>	<b>Yes</b>	5803	1015	85.11%
	<b>No</b>	377	1087	74.25%
	<b>Overall Percent</b>	74.62%	25.38%	83.19%
<b>Holdout</b>	<b>Yes</b>	2505	430	85.35%
	<b>No</b>	159	460	74.31%
	<b>Missing</b>	0	0	
	<b>Overall Percent</b>	74.54%	25.46%	83.43%

The classification table shows that nearest neighbor does fairly well at predicting the likelihood of death or otherwise in children under five years. Of the cases used to create the model, 5803 of the 6818 children alive were classified correctly. 1087 of the 1464 child deaths were classified correctly. Overall 83.2% of the training cases were classified correctly, corresponding to the 16.8% incorrect shown in the model summary table. The hold out sample also reveals that 2505 of the 2735 children alive were classified correctly while 460 of the 619 child deaths were classified correctly. Overall, 83.2% of the testing cases were correctly classified by the model. The overall classification for the holdout sample was 83.4%, which is very close to the testing samples. This suggests that, overall the model was about 85.1% correct classification for children alive while that of children who are not alive was about 74.3%.

## 4.6 Comparison of Models

Table 4.20: Comparison Table

Prediction	Percentage Correct (Child is Alive)		Overall Percentage	
	Yes	No		
<b>Logistic Regression</b>	83.9	67.2	81.0	
<b>Neural Network</b>	Training	82.3	65.4	79.4
	Testing	84.2	65.2	80.7
	Holdout	82.7	63.8	79.5
<b>Kth Nearest Neighbor</b>	Training	85.1	74.3	83.2
	Holdout	85.4	74.3	83.4

From Table 4.12 Nearest Neighbor dominated over the other techniques in terms of the overall percentage of the correct classification with an average of about 83.3%. Both Neural Network and Logistic Regression perform almost the same in the overall correct classification. Nearest Neighbor also performs better in classifying deaths correctly in children under five (74.3%) followed by Logistic Regression (67.2%) while Neural Network recorded the least.

In this study, Correct Classification Rate was used as the measure to compare the three different models. It is the number of correctly classified claims divided by the total number of claims in the data set.

$$\text{Correct Classification Rate} = (a + d)/N$$

The model with the highest Correct Classification Rate will be selected as the most appropriate model.

Table 4.21: Correct Classification rate Table

		Actual +	Actual -	CCR
<b>Logistic Regression</b>	<b>Predicted +</b>	8187	1566	<b>0.81</b>
	<b>Predicted -</b>	683	1400	
<b>Neural Network</b>	<b>Predicted +</b>	8070	1683	<b>0.80</b>
	<b>Predicted -</b>	726	1358	
<b>Kth Nearest Neighbor</b>	<b>Predicted +</b>	8303	1445	<b>0.83</b>
	<b>Predicted -</b>	536	1547	

The CCR table reveals that the k<sup>th</sup> Nearest Neighbor was the most efficient in modeling Mortality in Children under five with a CCR of 83%. This is followed by Logistic Regression with a CCR of 81% and the least was Neural Network with a CCR of 80%. It is evident that the difference in the CCR amongst the three models was quite insignificant. It can therefore be concluded that although Nearest Neighbor was the best predictor of mortality in children under five years, the other models will also do a good job in modeling mortality in children under five years in Ghana.

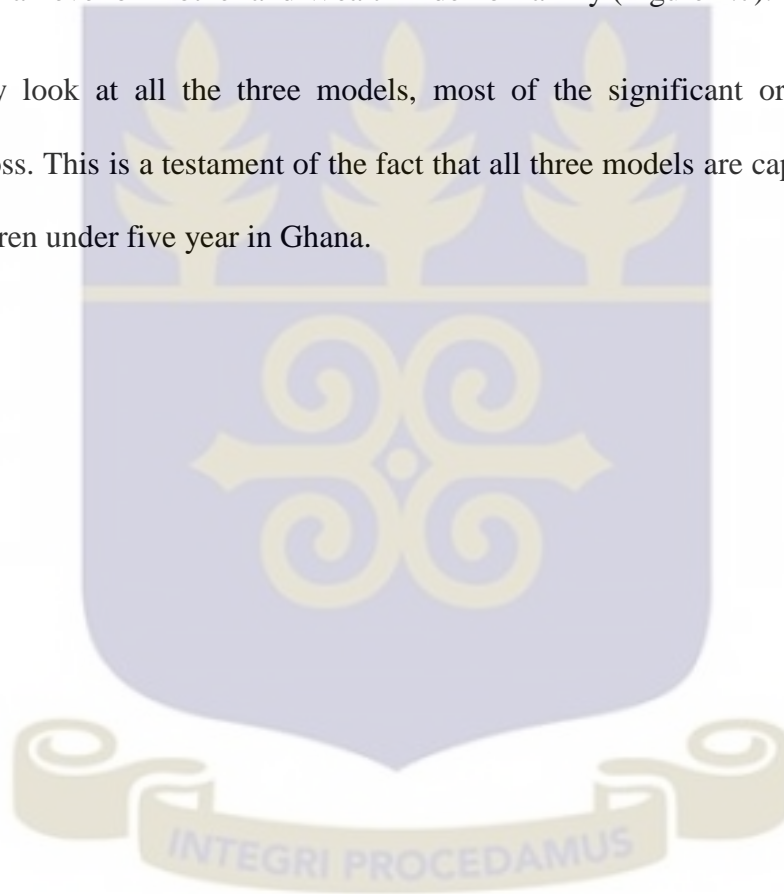
A comparison of the predictor variables that were significant or most important across all the three models would also be employed.

For Logistic Regression, Mothers' highest educational level, Age of mother at birth, Type of toilet facility used by family, Wealth index of family, whether mother is registered with health insurance, Sex of child and alcohol consumption were found to be significant in predicting mortality in children under five years (Table 4.8).

For the Neural Network model, Age of mother at birth was the most important variable in predicting mortality in children followed by Mothers' highest educational level, Type of toilet facility used by family, alcohol consumption and wealth index of family (Table 4.14).

For the Nearest Neighbor model, the most important contributor to the model was Type of toilet facility used by the family. This was followed by Age of mother at birth, alcohol consumption, Highest Educational level of Mother and Wealth Index of family (Figure 4.7).

Taking a cursory look at all the three models, most of the significant or highly important variables cut across. This is a testament of the fact that all three models are capable of predicting mortality in children under five year in Ghana.



## CHAPTER FIVE

### SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

#### 5.1 Introduction

This chapter deals with the summary of the findings as well as conclusions and recommendations for future research. The results and findings of the modelling techniques applied to the data will be presented.

#### 5.2.0 Summary

##### 5.2.1 Logistic Regression

In the logistic regression model, Mothers' educational level, Age of mother at birth, Type of toilet facility used, Wealth index of family, whether mother is registered with health insurance, Sex of child and alcohol consumption were found to be significant in predicting mortality amongst children under five in Ghana. The model predicted 81.00% over all prediction rate. The model also correctly classified the probability of a child dying at 67.21% and a child being alive at 83.94%.The model is therefore useful at predicting mortality in children.

##### 5.2.2 Neural Networks

In the Neural Network model, Mothers' educational level, Age of mother at birth, Type of toilet facility used, alcohol consumption and wealth index were found to be the most important variables in predicting mortality amongst children under five in Ghana.

The model predicted overall of 79.4% of the training cases, 80.7% of the testing cases and 79.5% holdout cases. The model also correctly classified the probability of child dying at 65.4% for training cases, 65.2% for testing cases and 63.8% for holdout cases. The model also correctly classified the probability of an alive child at 82.3% for training cases, 84.2% for testing cases and 82.7% for holdout cases. The area under the ROC curve revealed that, for a randomly selected child who is alive or a randomly selected child who is dead there is a 0.701 probability that the model-predicted pseudo-probability will be higher for a child who is alive than for a child who is dead. The model is also useful at predicting mortality in children.

### **5.2.3 K Nearest Neighbor**

In the K Nearest Neighbor model, Mothers' educational level, Age of mother at birth, Type of toilet facility used, alcohol consumption and wealth index were found to be the most important variables in predicting mortality amongst children under five in Ghana.

The model predicted overall of 83.2% for the training cases and 83.4% for the holdout cases. The model also correctly classified the probability of child dying at 74.3% for both training cases and holdout cases. The model also correctly classified the probability of an alive child at 85.1% for training cases and 85.4% for holdout cases. The model was found to be the most useful at predicting mortality in children with a CCR of 83% compared to Logistic Regression(81%) and Neural Network(80%).

### 5.3 Conclusion

The study sought to compare three different models in the prediction of mortality in children under five years. The three model were Logistic Regression, Neural Network and k Nearest Neighbor. The study revealed that the Nearest Neighbor was the best model for predicting mortality in children under five years. Logistic regression and Neural Network will also do a good job in predicting mortality in children.

The study also sought to identify the predictor variables that were most significant or important in the models. Taking a cursory look at the models, Logistic Regression (Table 4.8), Neural Network (Table 4.14) and Nearest Neighbor (Figure 4.7) five variables cut across as the most significant or important in predicting mortality in amongst children under five in Ghana. Mothers' highest educational level, Age of mother at birth, Type of toilet facility used, alcohol consumption and wealth index were found to be the most important variables in predicting mortality in amongst children under five in Ghana across all models.

For the highest level of education of mother, mothers with higher education experienced fewer child mortality cases compared to mothers with no education (Table 4.1). This result is confirmed by a previous research conducted by the institute of Health Metrics in the University of Washington in 2011 which revealed that Women with more education tend to have smaller families, in part because of increased employment opportunities and better knowledge about contraception; fewer children in a family increases the chances that an infant will survive. More education also helps women make better decisions about the health of their children.

Type of toilet facility used by family was another important variable across the models. There were fewer deaths recorded for families who use flush toilets as compared to those using pit

latrines or no toilet facility (Table 4.1). Various studies have established the effects of type of toilet facility on child mortality. According to a study conducted by Wak (2002) in the Kassena-Nankana district in Ghana, it was found that, child mortality is experienced by children whose compounds have no toilet facilities while those who use water closet or pan toilet latrine experience the lowest child deaths.

Wealth index of family was another significant variable. There were fewer deaths recorded for families who are rich as compared to those from poor families (Table 4.1). Generally, several studies have indicated that children born by mothers with poor household index are more likely to experience infant mortality than children born by mothers with rich household index (Mutanga,2004)

Alcohol consumption was also another significant variable in the models. Interestingly there were fewer deaths recorded for mothers who consumed alcohol as compared to those who did not consume alcohol (Table 4.1).

Age of mother was another significant variable across all models. Several studies have revealed that, the age of the mother is statistically significant in explaining infant and child mortality. It is generally expected that children born to young mothers (aged less than 20 years) and those born to older mothers (aged 40-49 years) should have higher mortality than those born to mothers aged 20-39 years (Mustafa &Odimegwu, 2008, Kembo&Ginneken, 2009).

#### 5.4 Recommendations

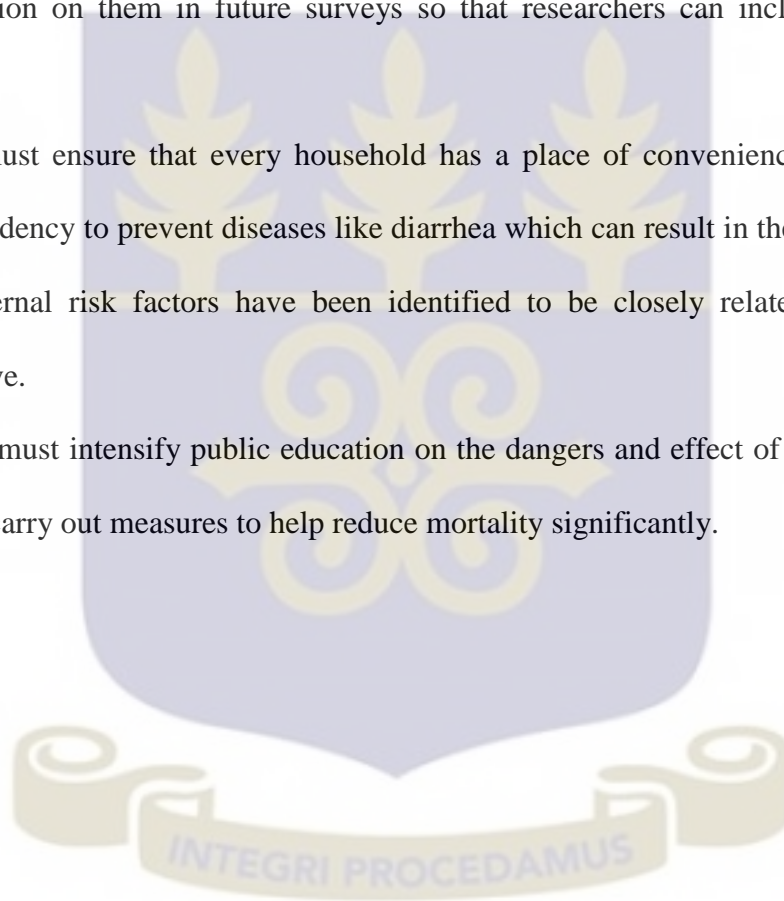
A future study on child mortality should consider the following:

Responses to very important variables like Size of child at birth, post natal care , pre natal care, and place of delivery in the GDHS data were very few and hence was not included in the study.

These are also very important variables and therefore all respondents should be encouraged to provide information on them in future surveys so that researchers can include them in their analysis.

Policy holders must ensure that every household has a place of convenience that is hygienic which has the tendency to prevent diseases like diarrhea which can result in the death of children under five. Maternal risk factors have been identified to be closely related to mortality in children under five.

The government must intensify public education on the dangers and effect of child mortality on society and also carry out measures to help reduce mortality significantly.



## REFERENCES

- Abedi K.(1991). *Factors influencing infant mortality in Vietnam*: Journal of Biosocial Science 1993; 25(3): 285-302.
- Adhikari R, Podhisita C(2010). *Household headship and child death; Evidence from Nepal*: BMC Complement Altern Med ; 10:13. *African Population Studies, Vol. 19, No. 1, 2004, pp. 1-11*
- Akazili J, Doctor HV, Abokyi L, Hodgson A, Phillips JF(2011). *Is there any relationship between antenatal care and place of delivery? Findings from rural northern Ghana*: African Journal of Health Sciences 2011, 18:1-2.
- an application to boosting. In: Computational Learning Theory: 7 Second European Conference,*
- Anderson (1984). Exponential smoothing: The state of the art. *Journal of Forecasting*, 4, 1–28.
- Arokiasamy (2004). *The determinants of infant and child mortality in Tanzania*.
- Arriaga and Hobbs (1982) “*Birth Intervals and Childhood Mortality in Rural Bangladesh*”: *Demography*, Vol. 27, No. 2, 1990, pp. 251-265. [doi:10.2307/2061452](https://doi.org/10.2307/2061452)
- Artificial Intelligence, San Mateo, CA: Morgan Kaufmann, 1022–1027.*
- Asparoukhov et al., (2003). *Generalized Linear Models: A Unified Approach*: Thousand Oaks, CA: Sage Publications.
- Awoonor-Williams JK, Bawah AA, Nyongator FK, Asuru R, Oduro A, Ofosu A and Phillips JF(2013): *The Ghana essential health interventions program: a plausibility trial of the impact of health systems strengthening on maternal & child survival*: BMC Health Serv Res, 13(Suppl 2):S3.
- Baird J.(1975). *How many child deaths can we prevent this year?*: Lancet ;362(9377): 65-71.

Baqui AH, Ahmed S, Arifeen EL. S, Darmstadt GL, Rosecrans AM, Mannan I(2009) *Effect of timing of first postnatal care home visit on neonatal mortality in Bangladesh: a observational cohort study*: BMJ; 339: b2826.15.

Baron A(1991). *Delayed breastfeeding initiation increases risk of neonatal mortality*. Pediatrics 2006; 117(3): e380-386.

Bayne C., Beauchamp R., Kane V.,McCabe D.(1983). *Gender imbalance in infant mortality: A*

Bayne E.,Saka F.,Futa G.(1993). *An Introduction to Generalized Linear Models*, 2 ed. Boca Raton, FL:

Bell J.,Ribar K.,Vechio Y.(1990)

Bellman, Lohse and Johnson, (1999). *Modern Factor Analysis*, 3rd ed. Chicago: University of Chicago Press

Bishop (1995). "*Mortality and Development Revisited*," *Quantitative Studies of Mortality Decline in the Developing World (World Bank Staff Working Papers Number 683, Population and Development Series Number 8)*, The World Bank, Washington DC,.

Black RE, Morris SS, Bryce J. (2003) *Where and why are 10 million children dying every year?* Lancet; 361(9376): 2226-2234.

Buckley G. (2003). "*Mortality and Development Revisited*," *Quantitative Studies of Mortality Decline in the Developing World (World Bank Staff Working Papers Number 683, Population and Development Series Number 8)*

Caldwell,Reddy(1983).The Bellagio Child Survival Study Group. How many child deaths can we prevent this year?

Chatterjee R.,Barcum O.(1970). An algorithm for finding best matches in logarithm expected time. *ACM Transactions on Mathematical Software*, 3, 209–226

Chiang, Zhang and Zhou (2006). *Simple models for the analysis of association in cross-classifications having ordered categories*. Journal of the American Statistical Association, 74, 537–552.

Cizerk and Fitzgerald(1999). *Multiple Regression in Practice*. BeverlyHills, CA

Claeson K. and Larrson R. (2003). *Factors affecting mortality in children in sub Saharan Africa*

Clark K. , Radcliff J.,Singha R.(2013):*Evidence-based, cost-effective interventions: how many newborn babies can we save?*the Lancet Neonatal Survival Steering Team

Clearly U. , Angel A.(1984).*Multivariate data analysis*. New York: John Wiley & Sons, Inc.

Cleland V.,Ginneken V.(1988). *The Stalling Child Mortality in Ghana: The Case of the three Northern Regions*. Paper Presented at the 5th Conference of Union for Africa Population Studies (UAPS) at Arusha, Tanzania, 10-14 Decemeber, 2007. Health Policy Planning 1995, 10(4):384-394.

Cochrane I.,Leslie J.,Ohara K.(1980). *Does Access to Water and Sanitation Affect Child Survival? A Five Country Analysis*. Background paper for Human Development Report. New York.

Countdown Coverage Writing Group (2008)

Crawley A.(1979). *Applied Econometrics for Health Economists*. A Practical Guide. 2nd edition, York, UK

Crawley E(1979). *Fast SVM training algorithm with decomposition on very large data sets*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27,603–618.

**cross-national study of social structure and female infanticide. *Social Science & Medicine* 62: 360-374.**

Currie and Moretti(2003). *"The Influence of Child Spacing on Child Survival,"* *Population Studies*, Vol. 38, No. 1, 1984, pp. 47-72.

Da Vanzo and Ranka K.(1993). *"Birth Order and Child Nutritional Status: Evidence from Malaysia,"* *Economic Development and Cultural Change*, Vol. 36, No. 2, 1988, pp. 341-354

Das gupta m. (1990). *"Death clustering, mother's education and the determinants of child mortality in rural Punjab, India"*, *Population Studies*, 44, 489-505. *Demography*, 20, 1, 1-26.

Davis,Edelman ,Gammerman(1992). *A decision theoretic generalization of on-line learning, an application to boosting*. In: *Computational Learning Theory: 7 Second European Conference, EuroCOLT '95*, , 23–37.

Degeratu, Rangaswamy and Wu, (2000). *Generalized Linear Models and Extension*. Station, TX: Stata Press

Deribew A, Tessema F, Girma B.(2007). *Determinants of under-five mortality in Gilgel Gibe Field Research Center, Southwest Ethiopia*. *Ethiopian Journal of Health Development* 2007; 21(2): 117-124.

Dey A., Astin O.(1993). *Regression models and life tables (with discussion)*. *Journal of the Royal Doctor HV*(2011). *Does living in a female-headed household lower child mortality? The case of rural Nigeria*. *Rural and Remote Health*11: 1635. (Online). Available: [www.rrh.org.au](http://www.rrh.org.au) (Accessed 23 January 2012).

Durand E.(1941). *Working set selection using the second order information for training SVM*. *Technical Report*. Taipei, Taiwan: Department of Computer Science, National Taiwan University.

Eisenbeis A., Avery Q(1972)

Ettarh RR, KimaniJ(2006). *Determinants of under-five mortality in rural and urban Kenya.EuroCOLT '95* , 23–37.

Finch K.,Schneider R.(2006).*An analytical framework for the study of child survival in developing countries*.Population and Development Review. 1984; 10(Suppl.): 24-45.

Fix ,Hodges(1952). *Bayesian network classifiers*.*Machine Learning*, 29, 131–163.

Ghana Statistical Service and Macro International Inc(1998): *Ghana Demographic and Health Survey 1998*. Accra, Ghana: Ghana Statistical service; 1999.

Ghana Statistical Service(2010). *Population and Housing Census, National Analytical Report*. 2013.(Accesson25thFebruary2014)[http://www.statsghana.gov.gh/docfiles/2010phc/National\\_Analytical\\_Report.pdf](http://www.statsghana.gov.gh/docfiles/2010phc/National_Analytical_Report.pdf) website

Grablowsky E.,Tallaey M.(1981).*Feedforward Neural Network Methodology*, 3rd ed. New York: Springer-Verlag

Grossman(1972). “*Socio-Economic Factors in Infant and Child Mortality: A Cross-National Comparison*,” Population Studies, Vol. 38, No. 2, 1994, pp. 193-223.

**GSS,GHS,ORC Macro (2008)**

Günther Fink<sup>1</sup>, Isabel Günther<sup>2</sup> and Kenneth Hill<sup>1</sup>(2007).*The effect of water and sanitation on child health: Evidence from the demographic and health surveys 1986–2007*

GyimahK.,Fernando R.(2002)*Maternal and Child Undernutrition Study Group: What works? Interventions for maternal and child under nutrition and survival*.

Han J. and Sun K.(2011).*Applications of Data mining Technique*,Population Studies, Vol. 55, No. 1, 2001, pp. 37-47. [doi:10.1080/00324720127679](https://doi.org/10.1080/00324720127679)

- Hand K., Henley E. (1997). *Series Evaluation of Tweedie Exponential Dispersion Model Densities*. *Statistics and Computing*, 15, 267–280
- Harrell F., Lee A. (1985). *Supervised and unsupervised discretization of continuous features*. In: *Proceedings of the Twelfth International Conference on Machine Learning*, Los Altos, CA: Morgan Kaufmann, 194–202
- Haykin (1998). *Applications Neural Networks. A Comparative Analysis*,” *Population Studies*, Vol. 39, No. 3, 1985, pp. 363-385. [doi:10.1080/0032472031000141576](https://doi.org/10.1080/0032472031000141576)
- Huberty R. (1994). *Perinatal mortality in Nigeria: do place of delivery and delivery assistants matter?* *Open Demography Journal* 1994; 4: 1-10.
- Johnson A., Wichern R. (2002). *Reproductive pattern, perinatal mortality, and sex preference in rural Tamil Nadu, south India: community based, cross sectional study*. *BMJ* 1997; 314: 1521-1524.
- Johnson G., Seshia V. (1992). *Improving regressor using boosting techniques*. In: *Proceedings of the 14th International Conference on Machine Learning*, D. H. Fisher, Jr., ed. San Mateo, CA: Morgan
- Jones R., Frederick K., Wristberg G. (2003). “*Infant and Child Mortality, Demographic and Health Surveys*, *Comparative Studies No. 20*,” Macro International, Calverton. Kaufmann, 107–115.
- Kembo R., Ginneken V. (2009) “*Women's empowerment in India and its states: Evidence from NFHS*”, *Economic and Political Weekly*, 39(7), 694–714. *Global Health An introductory textbook*. Lund: Studentlitteratur.
- Kembo, J., and Ginneken, K. V., (2009). *Determinants of Infant and Child Mortality in Zimbabwe: Results of multivariate hazard analysis*. Volume 21, Article

13, Pages 367-384 Published 01 October 2009

Kishor (2004). *Effects on Household Environmental Health Hazards to Child Survival in selected states in India*

Knoke V.(1982). k-Nearest Neighbor Classifiers. *Technical ReportUCD-CSI-2007-4, School of Computer Science and Informatics, University College Dublin,Ireland,*

Krzanowski(1975). *Elements of Continuous Multivariate Analysis*.Reading,MA: Addison-Wesley

Kubota L.,SamataN.,Jamba R.(2004). *Accelerating reproductive and child health programme impact with community-based services: the Navrongo experiment in Ghana*. PLoSOne , 8(3):e58924. doi:10.1371/journal.pone.0058924

Lancet(2007). “*How Bio- logical and Behavioral Influences on Mortality in Malaysia Vary during the First Year of Life,*” Population Studies, Vol. 37, No. 3, 1993, pp. 381-402.

Larose (2005). *Pattern Recognition Technologies, Metaanalysis*.IEG WorkingPaper. Washington, DC

Lin Q,HuangK.,Chan C.(2004). *Components of under-five mortality trends, current stagnation and future forecasting levels*.National Commission on Macroeconomics and Health Background papers-Burden of Disease in India.

Lindenbaum,Chakraborty,Elias(1985). *Determinants of Infant and Under-Five Mortality – The Case of Jordan*

Lungu I. and Bara A.(2012). *Data Mining Techniques, Applications in Medical Sciences,*

Meara (2001). “*Infant and Child Mortality Determinants in Bangladesh: Are They Changing?*” *Journal of Biosocial Science*,Vol. 29, No. 4, 1997, pp. 385-399. doi:10.1017/S0021932097003854

Meshbane L.,Moris V.(1996). *Applying an equity lens to child health and mortality: more of the same is not enough*. Lancet 2003; 362: 233-241.

Miller J.E., Trussell J., Pebley A.R., Vaughan B. (1992). "*Birth spacing and child mortality in Bangladesh and Philippines*", *Demography*, 29(2), 305-318.

Miller(1992). *Birth intervals, breastfeeding and determinants of childhood mortality in Malawi*. *Social Science & Medicine* 1999;48(3): 301-312.

Mishara R., Retherford R.(2013).*Determinants of utilisation of intrapartum obstetric care services in Cambodia, and gaps in coverage*. *Global Public Health* 2011, 6:8. 890-905 *Models*, M. I. Jordan, ed. Cambridge, MA: MIT Press, 301–354.

Mosley R., Chen K.(1984).*Socioeconomic factors differentiating maternal and child health-seeking behavior in rural Bangladesh: a cross-sectional analysis*. *Int J Equity Health*, 9:9.

Murthi Y. and DrezeK(2001). "*Socio-Economic Factors in Infant and Child Mortality: A Cross-National Comparison*," *Population Studies*, Vol. 38, No. 2, 1984, pp. 193-223

Mustafa, H. E, and Odimegwu, C, (2003): *Socioeconomic Determinants of Infant Mortality in Kenya: Analysis of Kenya DHS 2003*. *Journal of Humanities and Social Sciences* ISSN 1934-7227, Volume 2, Issue 2, 2008.

Mutanga(2007). *Infant Mortality and Mother's Education in Ondo State, Nigeria*. *Social Science and Medicine*; 40(2):253-263.

Mutunga (2004).*Some Socio-economic and Demographic Determinants of infant and child mortality in Tanzania: A case study of Karagwe District, KageraRegion*.*Open Demography Journal* 2011, 4:1-10.

Myers V.,Forgy Y.(1963). *Multi-interval discretization of continuous-value attributes for classification learning*. In: *Proceedings of the Thirteenth International Joint Conference*

- Olmez ,Dokur (2010). *Under-Five Mortality in Lesotho*:(Unpublished Thesis RIPS- University of Ghana)
- Press V.,Wilson D.(1978).*BMD Biomedical computer programs*. Los Angeles: University of California
- Preston(1995). “Maternal Nutrition, Breastfeeding and Infant Survival,” In: W. H. Mosley, Ed., *Nutrition and Human Reproduction*, Plenum Press, New York,.
- Qiu J., Yung F., Sung R.(2014). *A hazards-model analysis of the covariates of infant and child mortality in Sri Lanka*.
- Rahman S, Salameh K, Bener A and El Ansari W(2010). *Socioeconomic associations of improved maternal, neonatal, and perinatal survival in Qatar*.Int J Womens Health 2010, 2:311-318.
- Raudys (2001).*Why do families of sick newborns accept hospital care?*A community-based cohort study in Karachi, Pakistan.
- Remus A.,Wong B.(1982). *Birth intervals, breastfeeding and determinants of childhood mortality in Malawi*. Social Science & Medicine 1999;48(3): 301-312Revised Edition, Comparative Studies No. 43, International Statistical Institute, Voorburg,
- Ridder, G., &Tunali, D. (2011).*Stratified partial likelihood estimation*. Journal of Econometrics, 92, 2, 193-232. Rural Remote Health 11:1635. Online). Available: <http://www.rrh.org.au> website
- Rosenberg K.,Gleit X.(1994). *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: SAGE Publications, Inc..
- Rustein(2000) "*On the Trail of 'Missing' Indian Females*", Economic and Political Weekly, 37(51), 5105-5118

Salari A.,Mung G.,Ranga U.(2013). Factors that affect under-five mortality among South African children: analysis of the South African Demographic and Health Survey Data Set.

Sen(1988).*Perceived barriers to utilization of maternal health services in rural Cambodia.* Health Policy, 95(2 3):255-263.

Shavlik et al.,(1991). *A Tutorial on Learning with Bayesian Networks.* In: Learning in Graphical Models, Mooney and Towell (1991). *Neural Networks: A Comprehensive Foundation*, 2nd ed. New York: Macmillan College Publishing.

UNICEF (2004).The state of the world's children. UNICEF, New York. NY. USA.

UNICEF, WHO, World Bank (2013), World Development Indicators

UNICEF, WHO, World Bank and UN-DESA Population Division(2014): *Levels and trends in childmortality.*2013.[http://www.who.int/maternal\\_child\\_adolescent/documents/levels\\_trends\\_child\\_mortality\\_2013.pdf?ua=1](http://www.who.int/maternal_child_adolescent/documents/levels_trends_child_mortality_2013.pdf?ua=1) website.

United Nations. The Millennium Development Goals Report(2010). NewYork: UN, 2010; 1-77

Wak, G.P, (2002).*Determinants of infant Mortality in the Kassena-Nankana District, Ghana.* (Unpublished Thesis RIPS -University of Ghana).

Weitof R.,Lasma B.,Ottang Y.(2003).Infant illness spanning the antenatal to early neonatal continuum in rural northern Ghana: local perceptions, beliefs and practices.

Welaga K.,Statiga R.,Kuntu U.(2013). “*Infant and Child Mortality in Tanzania and Ghana, Demographic and Health Surveys, Comparative Studies No. 20,*” Macro International, Calvert  
**WHO(2014)**

Wiginton R.(1980). Generalized collinearity diagnostics.*Journal of the American Statistical Association*, 87, 178–183.

Williams K.,LeeJ.,Fisher R.,Dickerman U.(1999). *Birth spacing, breastfeeding, and early child mortality in a traditional Indian society: a hazards model analysis*. Social Biology 1994; 41(3-4): 168-180.

Wilson A.,Sharda G.(2005). *Where and why are 10 million children dying every year?* Lancet 2003; 361(9376): 2226-2234

Wilson J.,Handgrave X.(1995). *Type of delivery attendant, place of delivery and risk of early neonatal mortality: analyses of the 1994-2007 Indonesia Demographic and Health Surveys*.

Wilson M.,Sharda S.(1994)

World Health Organization(2006).*Neonatal and Perinatal Mortality: Country, Regional, and Global Estimates*.

YarnodN.,Hart O.,Soltysik J.(1994). *4 million neonatal deaths: When? Where? Why?* Lancet 2005; 365: 891-900.

You D, Wardlaw T, Salama P and Jones G (2010).*Levels and trends in under-5 mortality, 1990–2008*. Lancet 2010, 375:100-103.

Zeng C.,Ching R.,Racus E.(2007). *Factors Associated With Trends in Infant and Child Mortality in Developing Countries during the 1990s*. Bulletin of the World Health Organization 78: 1256-1270

## APPENDIX

GET

FILE='C:\Users\Ansah\Desktop\ANALYSIS\GHBR5HFL - Copy - Copy\_1 - Copy.SAV'.  
DATASET NAME DataSet1 WINDOW=FRONT.

```

LOGISTIC REGRESSION VARIABLES Child_is_alive
/METHOD=ENTER PLACE_OF_RESIDENCE EDUCATIONAL_LEVEL
Source_of_drinking_water Toilet_facility Wealth_index Age_of_mother health_insurance Marital_status Sex_of_child alcohol_consumption
/CONTRAST (PLACE_OF_RESIDENCE)=Indicator
/CONTRAST (EDUCATIONAL_LEVEL)=Indicator
/CONTRAST (Source_of_drinking_water)=Indicator
/CONTRAST (Toilet_facility)=Indicator
/CONTRAST (Wealth_index)=Indicator
/CONTRAST (health_insurance)=Indicator
/CONTRAST (Marital_status)=Indicator
/CONTRAST (Sex_of_child)=Indicator
/CONTRAST (alcohol_consumption)=Indicator
/SAVE=PRED PGROUP COOK LEVER DFBETA RESID LRESID SRESID ZRESID DEV
/CLASSPLOT
/CASEWISE OUTLIER(2)
/PRINT=GOODFIT CORR ITER(1) CI(95)
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.85).
    
```

## Logistic Regression

### Notes

Output Created	22-FEB-2015 04:20:54
Comments	
Data	C:\Users\Ansah\Desktop\ANALYSIS\GHBR5HFL - Copy - Copy_1 - Copy.SAV
Active Dataset	DataSet1
Filter	<none>
Weight	<none>
Split File	<none>
N of Rows in Working Data File	11888

Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing
------------------------	-----------------------	--

Categorical Variables Codings					
		Frequency	Parameter coding		
			(1)	(2)	(3)
Highest educational level	No education	5087	1.000	.000	.000
	Primary	2662	.000	1.000	.000
	Secondary	3894	.000	.000	1.000
	Higher	193	.000	.000	.000
Wealth index	Rich	3370	1.000	.000	
	Middle	2207	.000	1.000	
	Poor	6259	.000	.000	
Marital status	Never married	211	1.000	.000	
	Currently married	10235	.000	1.000	
	Formerly married	1390	.000	.000	
Type of toilet facility	FLUSH TOILET	863	1.000	.000	
	PIT LATERINE	6719	.000	1.000	
	NO FACILITY	4254	.000	.000	
Source of drinking water	PIPED WATER	3819	1.000		
	NON PIPED WATER	8017	.000		
Alcohol consumption	No	8882	1.000		
	Yes	2954	.000		
Sex of child	Male	5967	1.000		
	Female	5869	.000		
Covered by health insurance	No	7061	1.000		
	Yes	4775	.000		
Type of place of residence	Urban	3908	1.000		
	Rural	7928	.000		

**Iteration History<sup>a,b,c</sup>**

Iteration	-2 Log likelihood	Coefficients
-----------	-------------------	--------------

		Constant
Step 0	1	11124.557
	2	11014.061
	3	11013.536
	4	11013.536

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 11013.536

c. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

**Correlation Matrix**

	Constant	PLACE_OF_RES IDENCE(1)	EDUCATIONAL _LEVEL(1)	EDUCATIONAL _LEVEL(2)
Constant	1.000	.015	-.802	-.799
PLACE_OF_RESIDENCE(1)	.015	1.000	-.025	-.022
EDUCATIONAL_LEVEL(1)	-.802	-.025	1.000	.963
EDUCATIONAL_LEVEL(2)	-.799	-.022	.963	1.000
EDUCATIONAL_LEVEL(3)	-.791	-.025	.960	.958
Step 1 Source_of_drinking_water(1)	-.022	-.291	.025	.014
Toilet_facility(1)	-.162	-.070	.186	.156
Toilet_facility(2)	-.106	.009	.053	-.005
Wealth_index(1)	-.095	-.448	.100	.080
Wealth_index(2)	-.049	-.242	.033	.021
Age_of_mother	-.487	-.005	-.001	.031

health_insurance(1)	-.112	-.004	-.041	-.043
Marital_status(1)	-.169	-.022	.011	.009
Marital_status(2)	-.308	.013	.015	.019
Sex_of_child(1)	-.107	.006	.012	.012
alcohol_consumption(1)	-.154	-.047	.007	.005

**Correlation Matrix**

	EDUCATIONAL _LEVEL(3)	Source_of_drinkin g_water(1)	Toilet_facility(1)	Toilet_facility(2)
Constant	-.791	-.022	-.162	-.106
PLACE_OF_RESIDENCE(1)	-.025	-.291	-.070	.009
EDUCATIONAL_LEVEL(1)	.960	.025	.186	.053
EDUCATIONAL_LEVEL(2)	.958	.014	.156	-.005
EDUCATIONAL_LEVEL(3)	1.000	.021	.132	-.017
Source_of_drinking_water(1)	.021	1.000	.000	-.065
Toilet_facility(1)	.132	.000	1.000	.476
Toilet_facility(2)	-.017	-.065	.476	1.000
Wealth_index(1)	.047	-.229	-.298	-.187
Wealth_index(2)	.002	-.211	-.088	-.204
Age_of_mother	.022	-.029	-.037	-.011
health_insurance(1)	-.021	-.034	-.029	-.018
Marital_status(1)	-.003	.002	-.011	.021
Marital_status(2)	.022	.041	.017	.049

Step 1

Sex_of_child(1)	.012	.002	-.008	.001
alcohol_consumption(1)	-.004	.000	-.056	-.168

**Notes**

Output Created	24-FEB-2015 15:36:37
Comments	
Data	C:\Users\salifu.nanga.WIUC-GHANA\Desktop\ANALYSIS\GHBR5HFL - Copy - Copy_1 - Copy.SAV
Input	
Active Dataset	DataSet1
Filter	<none>
Weight	<none>
Split File	<none>
N of Rows in Working Data	11888
File	

		Child is alive?			
		Yes		No	
		Count	Subtable Valid N %	Count	Subtable Valid N %
Type of place of residence	Urban	3253	27.4%	665	5.6%
	Rural	6547	55.1%	1423	12.0%
Highest educational level	No education	4192	35.3%	920	7.7%
	Primary	2126	17.9%	538	4.5%
	Secondary	3301	27.8%	606	5.1%
Source of drinking water	Higher	172	1.4%	23	0.2%
	PIPED WATER	3132	26.3%	692	5.8%
	NON PIPED WATER	6667	56.1%	1396	11.7%
Type of toilet facility	FLUSH TOILET	717	6.0%	148	1.2%
	PIT LATERINE	5435	45.8%	1300	11.0%
Covered by health insurance	NO FACILITY	3631	30.6%	638	5.4%
	No	5758	48.5%	1319	11.1%
Marital status	Yes	4026	33.9%	767	6.5%
	Never married	187	1.6%	24	0.2%

	Currently married	8496	71.5%	1784	15.0%
	Formerly married	1117	9.4%	280	2.4%
Sex of child	Male	4871	41.0%	1124	9.5%
	Female	4929	41.5%	964	8.1%
Alcohol consumption	No	7299	61.4%	1630	13.7%
	Yes	2496	21.0%	458	3.9%
Wealth index	Rich	2869	24.1%	516	4.3%
	Middle	1805	15.2%	406	3.4%
	Poor	5126	43.1%	1166	9.8%

GET

```
FILE='C:\Users\Ansah\Desktop\ANALYSIS\GHBR5HFL - Copy - Copy_1 - Copy.SAV'.
DATASET NAME DataSet1 WINDOW=FRONT.
*Multilayer Perceptron Network.
MLP Child_is_alive (MLEVEL=N) BY PLACE_OF_RESIDENCE EDUCATIONAL_LEVEL
Source_of_drinking_waterToilet_facilityWealth_indexhealth_insuranceMarital_statusSex_of_childalcohol_consumption WITH Age_of_mother
/RESCALE COVARIATE=STANDARDIZED
/PARTITION TRAINING=7 TESTING=2 HOLDOUT=1
/ARCHITECTURE AUTOMATIC=YES (MINUNITS=1 MAXUNITS=50)
/CRITERIA TRAINING=BATCH OPTIMIZATION=SCALEDCONJUGATE LAMBDAINITIAL=0.000005
SIGMAINITIAL=0.00005 INTERVALCENTER=0 INTERVALOFFSET=0.85 MEMSIZE=1000
/PRINT CPS NETWORKINFO SUMMARY CLASSIFICATION
/PLOT NETWORK ROC GAIN LIFT PREDICTED
/STOPPINGRULES ERRORSTEPS= 1 (DATA=AUTO) TRAININGTIMER=ON (MAXTIME=15)
MAXEPOCHS=AUTO ERRORCHANGE=1.0E-4 ERRORRATIO=0.0010
/MISSING USERMISSING=EXCLUDE .
```

## Multilayer Perceptron

Notes	
Output Created	21-FEB-2015 05:55:32
Comments	

Input	Data	C:\Users\Ansah\Desktop\ANALYSIS\GHBR5HFL - Copy - Copy_1 - Copy.SAV
	Active Dataset	DataSet1
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	11888
Missing Value Handling	Definition of Missing	User- and system-missing values are treated as missing.
	Cases Used	Statistics are based on cases with valid data for all variables used by the procedure.
Weight Handling		not applicable

Network Information			
Input Layer	Factors	1	Type of place of residence
		2	Highest educational level
		3	Source of drinking water
		4	Type of toilet facility
		5	Wealth index
		6	Covered by health insurance

		7	Marital status
		8	Sex of child
		9	Alcohol consumption
	Covariates	1	Age of mother
	Number of Units <sup>a</sup>		24
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1 <sup>a</sup>		7
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Child is alive?
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

**Model Summary**

Training	Cross Entropy Error	3831.846
	Percent Incorrect Predictions	17.5%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error <sup>a</sup>
	Training Time	0:00:00.58
Testing	Cross Entropy Error	1141.422
	Percent Incorrect Predictions	18.4%

Holdout	Percent Incorrect Predictions	17.0%
---------	-------------------------------	-------

