UNIVERSITY OF NAIROBI

COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES

SCHOOL OF MATHEMATICS

MODELLING THE FACTORS CONTRIBUTING TO UNDER-FIVE MORTALITY IN SOMALIA

By

ABDIRAHMAN OMAR DAHIR

A Research project submitted to the school of Mathematics, University

Of Nairobi, in Partial Fulfillment for the requirements of the Award of

the Degree of Master of Science in Social Statistics

June, 2015

Declaration

Declaration by the student

This research project is my original work and has not been submitted for a degree in any other university. Furthermore, the works by other authors has been duly acknowledged.

ABDIRAHMAN OMAR DAHIR

I56/68656/2013

....
Signature Date

Declaration by the supervisors

This project has been submitted for examination with my approval as University supervisor.

DR. NELSON O. OWUOR

....
Signature Date

Acknowledgements

Praise be to Allah for giving me the knowledge, wisdom, strength, sound mind to pursue my Master program in Social statistics because without His grace and support, I would not have succeeded.

I would also like to pass my sincere gratitudes to my supervisor, Dr. Nelson Owuor, for his scholarly advice, guidance, knowledge and skills that enabled me conduct and write this research project. I also appreciate the support from the school of mathematics, University of Nairobi, more specifically, the lecturers who molded my mind to be able to conduct the research project.

I also pass my utmost gratitude to my parents, brothers and sisters for their moral support, and most important, their prayers that have seen me through my graduate studies.

I also thank my dear brother Fahad Yasin Dahir for his financial support for the entire masters program; thank you my brother.

Last but not least, my appreciation goes to my classmates for their encouragement and comments: especially Arthur Yambayamba, Madalina kaku, Johnson Mwangi.

Dedication

This research project is dedicated to my parents and my brother

Fahad Yasin Dahir

Abbreviations and Symbol

Symbols

$J(V)$: Score function

$\sum_{i=1}^{n} Z_i$: Summation of Z_i

$\overline{Z}$: Mean of Z

V^T: Transpose of V

S_B: Between scatter matrix

S_W: Within scatter matrix

S^{-1}: Inverse of S

η: Canonical correlation

$\wedge$: Wilks lambda

λ: Eigenvalue

χ^2:   Chi-square

df: degrees of freedom

H_0:   Null hypothesis

C: Optimal cutting score

C_{PRO}:   proportional chance

Abbreviations

MDGs: Millennium Development Goals

MICS :   Multiple Indicator Cluster Survey

UNICEF: United Nations Children's Fund

UNDP :   United Nations Development Programme

WHO :   World Health Organization

PAPFAM: Pan Arab Project for Family

ECA :   Economic Commission for Africa

CSZ :   Central South Zone

U5MR :   Under Five Mortality Rate

LDA :   Linear Discriminant Analysis

MCC :   Maximum Chance Criteria

PPS :   Probability Proportional Size

Abstract

Under-five mortality rate is the key indicator of both child well-being and coverage of child survival interventions factoring social and economic development. This is in line with Millennium Development Goal 4 (MDG 4) projects reduction of under-five mortality rate by two-thirds by 2015. Somalia is one of the countries with the highest mortality rate in the world. This study was conducted to identify the factors contributing to under-five mortality in Somalia using discriminant analysis. The data used was from UNICEF 2006 Multiple Indicator Cluster Survey (MICS). Using discriminant analysis, a stepwise procedure was used to identify only the significant variables which were ranked according to Wilk's Lambda values. The canonical discriminant function coefficients (unstandardized and standardized) were also calculated for independent variables. Based on this procedure, children ever born, source of drinking water, age of the mother, current marital status of the mother and region of residence were found to be significantly contributing to under-five mortality in Somalia. The classification accuracy of the model was 73.8%. Therefore, the discriminant function constructed was adequate and thus can be used to classify a child into any of the two groups, dead or alive, based on significant factors that are contributing to under-five mortality.

Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

1.1 Introduction

Under-five mortality (U5MR) is the probability of dying between birth and the fifth birthday. It is usually expressed as number of deaths per 1,000 live births, [20].

Under-five mortality rate is a key indicator of both child well-being, and the coverage of child survival interventions factoring in social and economic development. In line with this, Millennium Development Goal 4 (MDG 4) projects reduction under-five mortality rate by two-thirds by 2015. Consequently, substantial progress has been made by reducing the rate from 90 to 46 deaths per 1,000 live births between 1990 and 2013 respectively; which accounts for 49 percent reduction. The world has witnessed accelerated under-five mortality reduction rate in the last two decades than before. The global annual rate has in effect reduced three times between 1990 and 1995. However, to achieve MDG 4 on time, the global annual rate of reduction in under-five mortality ought to rise to 20.8 percent for 2013-2015,

much higher than the 4.0 percent achieved between 2005 and 2013.

Comparatively, the under-five mortality rate in developing countries was 76 deaths per 1,000 live births more than 12 times the average rate in developed countries. As such, many countries still have very high rates. Most of these countries are in Sub-Saharan Africa, comprising all 12 countries with an under-five mortality rate of 100 or more deaths per 1,000 live births, [13].

Somalia exhibits one of the highest infant mortality rates in the world. UNICEF (2006) estimates the under-five child mortality rate (U-5MR) at 224 per 1,000 (which implies one in every five children die before reaching the age of five). The Central South Zone (CSZ) experiences the highest Under-5 mortality rates. In 2006, it was estimated at 144 death per 1000 live births, and reduced significantly from 224 deaths per 1000, [20].

1.2 Background

Worldwide under-five mortality rate has declined from an estimated 90 deaths per 1000 live births in 1990 to 46 deaths per 1000 live births in 2013. Sub-saharan Africa, however, remains one the regions with highest figures estimated at 92 deaths per 1000 live births, more than 15 times the average for developed regions, [20] . According to United Nations Development Programme (UNDP) research, the under 5 deaths in Sub-Saharan Africa were 3,709 in 1990 and reduced to 3,019 in the year 2013. The reduction of risks of dying before age 5: for instance, during 2000-2005, under-five mortality stood at 9 per 1000 in the more developed regions but at 153 per thousand in the least developed countries (United Nations, 2007). It is noted that the gap between more developed and the less developed regions in the world is larger in proportional terms for death rates in early childhood than for those in adult ages. Various researches prove that under-five mortality levels are influenced mainly by poverty, education, particularly, of the mothers; availability, accessibility and quality of health services; environmental risks including access to safe water and sanitation, and nutrition.

Somalia is one of the countries with highest Under-five mortality rates, estimated at 180 deaths or more per 1,000 live births, [13]. There is need for accelerated efforts if the MDGs target is to be attained.

In Somalia under-five mortality rate is among the highest in the world. Nearly one in every 12 Somali children dies before reaching age one, while one in every 7 does not survive to the fifth birthday [20]. The last three decades of armed conflicts, lack of functioning government, economic collapse, and disintegration of the health system and other public services - together with recurrent droughts and famines has turned Somalia into one of the worlds most difficult environments for survival, [22]. The absolute number of under-five mortality rate has been on an

3

increase since 1990 due to combination of high population growth and nearly static resource allocation, [5].

Albeit estimates showing decrease in under-five mortality rate in Somalia, the country still hosts the worst infant, child and maternal mortality rates. The highest under-five mortality rate is in the South Central zone, with significant variability across the zones. Between 1999 and 2006, good progress has been made in decreasing the under-five mortality rate across all three zones of Somalia, [19]. The 2006 data set for infant mortality indicated that efforts to decrease infant mortality yielded some positive results, slightly outperforming the set target, that is, between 1990 and 2006, infant mortality rate decreased from 152 deaths per 1,000 live births to 86 deaths per 1,000 live births. The South Central zone, in 1999, notably had the highest infant mortality rate amongst the zones and by 2006, Somalilands infant mortality had exceeded that of Puntland and South Central despite all three zones having similar rates for 2006. Several challenges in reducing child mortality rates are still faced by Somalia in spite of having surpassed the MDG indicators. A number of recommendations aimed at improving the progress towards this MDG for Somalia have been proposed by the UNDP (2012):

- Improve access to and quality of health facilities in rural areas, with a greater emphasis on post-natal care by expanding basic infrastructure such as roads, water and electricity, timely supply of medicines, employing qualified staff and building more health facilities.

- Sensitize and encourage the use of mosquito nets especially for children under five.

- Undermine malnutrition by launching programmes aimed at raising awareness among mothers on health, hygiene and maternal feeding practices.

- Involve the government in leading the provision of improved health services for the marginalized communities.

1.2.1 Under-five Mortality Rate in Somalia

Somalia is the poorest country in Sub-Saharan Africa and one of the poorest countries in the world. According to the World Bank social economic survey 2002, Somalia has not appeared in the human development global rankings over the last 10 years due to lack of data and the absence of a central government.

The mortality ratio is very high at a rate of 1,044 per 100,000 live births. There are high illiteracy levels in the country and this can be explained by the fact that only 18% and 21% of the girls and boys respectively get the basic education at the primary school level.

On a positive note, more than 21% of the population has access to clean water from improved water sources.

For our case study, we will look at the findings of a research done previously and compare the results with the data analysis in this paper. The 2006 Multiple Indicator Cluster Survey (MICS) is a nationally representative survey of 5969 households, 6764 women age 15-49 and 6305 mother's and caretakers of children age less than five. The primary purpose of the MICS is to provide policy makers, shareholders and planners with reliable and detailed information needed to monitor the situation of women and children in Somalia. Information on child mortality, sex of the infant, water and sanitation, mothers' education and marital status of the mother is included.

Child Mortality

As per 2006 survey on mortality levels, one in every twelve Somali children dies before reaching a year old, while one in every 7 does not survive to the fifth birthday. The survey further shows that the highest levels of mortality are found in the Central South Zone.

Sex of infant

Moreover, the survey showed that male children experience higher mortality than

female children. The sex difference is especially pronounced for infant mortality.

Sanitation

Surprisingly, half of the Somali population is living without any type of toilet facilities. Further, 37% are using a facility with sanitary means of extra disposal. Slightly over three quarters of Somalis living in urban areas are using sanitary means of extra disposal compared to 13% of people living in rural areas. When it comes to disposing of child's waste, over a third of children age 0-2 months (35%) have their stools disposed of in a safe way.

Literacy

The survey showed that quarter of Somali women aged between 15-24 are literate, while, those living in urban areas are four and a half times more likely to be able to read than those living in urban areas.

Marriage

Eight percent of women aged 15-49 years were married by the time they were 15. The proportion increases to 46 percent by the time the women are 18. Furthermore, a quarter of Somali women aged 15-19 were married by 2006. In thirty-one percent of these marriages, the husband is ten years older than the woman. Twenty-three percent of women married in 2006 are married to men who are in a polygamous union. Older women and those with no education are more likely to be in a polygamous union and vice-versa.

Below is a table showing childhood mortality rates in Somalia as per the Somali MICS/PAPFAM 2006.

Table 1.1: childhood mortality rates in Somalia

Years preceding Survey	Neonatal Mortality	Post Neonatal Mortality	Infant Mortality	Child Mortality	Underfive Mortality
0 - 4	41	45	86	53	135
5 - 9	37	54	91	67	152
10 - 14	50	59	109	94	194

Source: UNICEF, 2006

Table 1.2: Child mortality by sex and residence

Years preceding Survey	Neonatal Mortality	Post Neonatal Mortality	Infant Mortality	Child Mortality	Underfive Mortality
Sex					
Male	43	48	91	53	139
Female	33	43	76	54	126
Zone					
North West	36	52	88	27	113
North East	35	45	80	46	122
Central South	44	43	87	66	144
Residence					
Rural	40	48	88	50	134
Urban	42	44	85	55	135

Source: UNICEF, 2006

Table 1 presents neonatal, post neonatal, infant, child and under-five mortality rates for the three recent five year periods before the survey. Neonatal mortality in the most recent period is 41 per 1000 live births. This rate is similar to post neonatal deaths (45 per 1000 live births) during the same period; that is, the risk

of dying for any Somali child who survived the first month of life is similar as in the remaining 11 months of the first year of life. Hence, just less than 50 percent of infant deaths in Somalia occur during the first month of life.

The infant mortality rate in the five years preceding the survey is 86 per 1,000 live births and under-five mortality is 135 deaths per 1,000 live births for the same period. This means that one in every 12 Somali children dies before reaching age one, while one in every 7 does not survive to the fifth birthday.

Mortality trends can be examined in two ways: by comparing mortality rates for five year periods preceding a single survey and by comparing mortality estimates obtained from various surveys. However, these comparisons should be interpreted with caution because quality of data, time references and sample coverage varies. In particular, sampling errors associated with mortality estimates are large and should be taken into account when examining trends between the survey.

1.3   Problem statement

Somalia like all other countries in the world is committed to attaining the targets embodied in the MDGs by the year 2015. However, due to political instability there is little understanding whether Somalia will be capable of achieving all the MDGs, in particular reducing under-five mortality. It is therefore, necessary to identify the factors contributing to under-five mortality.

1.4   Objectives of the study

The overall objective of this study was to model the factors contributing to under-five mortality in Somalia.

The specific objectives were:

1. To identify socio-economic, bio-demographic and environmental factors that are significantly associated with under-five mortality using stepwise discriminant procedure.

2. To determine the classification accuracy of the discriminant model resulting from the classification table of the holdout sample.

1.5 Key Research Questions

This research was guided by the following research questions:

1. What are the factors that are significantly associated with under-five mortality?

2. How accurate is the discriminant model constructed in terms of making classification?

1.6 Justification of the study

Policy makers and other stakeholders conducting programs aimed at reducing under-five mortality need analytical tools for them to make informed decisions. A statistical model can assist in making such decisions both in the present and future based on predictions that can be made.

Finally, other researchers often rely on existing information in order to make further improvements and deepen the understanding of the subject matter under-five mortality in the case of this study.

CHAPTER 2

LITERATURE REVIEW

2.1   Introduction

In this chapter, we review how different scholars and researchers have used various statistical models and methods to determine the factors that contribute to under-five mortality in various regions of the world. In particular, the section focuses on the objectives of the studies, the methods of data collection, data analysis approaches, variables used in the study and the findings and conclusion as per the underlying analysis. Further the topic covers an illustration of the conceptual framework and finally a summary of the review.

2.2   Literature Review

Yiadom et al. in 2014 studied the analysis of under-five mortality in Ghana using logit regression model. The study sort to identify the significant determinants of under-five mortality in the Tano South district of Ghana. The target popu-

lation was mothers where they were to state whether or not they experienced under-five mortality. 200 mothers were investigated, however, the data collected was also supplemented with information on environmental and clinical factors. The analysis showed that grand-multigravida parity, anemia, and Malaria were significantly associated with increased odds while the use of treated bed-nets, child vaccination, practice of exclusive and not-exclusive breast feeding were significantly associated with decreased odds of under-five mortality. The study recommended for more awareness on the benefits of using treated bed nets, child vaccine, and practice of exclusive and not-exclusive breast feeding by residents of Tano district [3].

Geneti and Deressa in 2014 examined the Under-age five children mortality (U5CM) as a population health determinant and as an indicator of the social and economic development of a population. The study sort to statistically determine the correlates of child mortality in Ethiopia. The data used was EHDS data of 2011. From the analysis, the overall prevalence of mortality among children in Ethiopia was 11.3%. To model the effects of selected socio-economic, demographic, health and environmental predictors, logistic regression was used. From the analysis, the sex of the child, preceding birth interval, birth order of the child, place of residence, mother's education level, available toilet facility, number of house members, and mothers' age at birth and source of drinking water were the most significant determinants of children mortality in Ethiopia. The study recommended for female literacy programs in the country as a policy to reduce child mortality. Furthermore, the women needed to be enlightened on the risk of early marriage and its effect to child birth. Finally, it was recommended that the government should encourage household sanitation program through extensions workers to reduce child mortality,[7].

Chowdhury in 2013 examined determinants of under-five mortality in Bangladesh.

The data used was obtained from Bangladesh demographic and health survey done in 2007. The study used Chi-square test for independence and multivariate proportional hazard analysis. It was found that, father's education, place of residence, region of residence, number of children under five years of age, previous death of sibling, mother's age and breastfeeding had a significant effect on under-five mortality. Further, the proximate determinants were found to have a stronger influence on under-five mortality as compared to the socioeconomic factors. The study recommended for the need to increase paternal care, provide more civic facilities in vast rural areas and motivate couples on the use of modern contraceptive methods for longer birth periods so as to reduce childhood mortality in Bangladesh. It also advocated for persuasion of mothers for full breast feeding practices, [4].

Mani in 2012 studied the effects of programmable determinants on under-five mortality through accounting for family-level clustering and adjusting for background variables by use of Cox frailty model and Cox proportional hazards models. The Cox frailty model indicated that the mother's age at birth, place of delivery, sex of the baby, composite of birth order and birth interval, baby size at birth and breastfeeding had a significant effect on the under-five mortality. Further, it was showed that the in richest quintile, children had significantly lower mortality as compared to the poor quintile. The determination of hazard ratios for determinants showed same results for all the three models except the death of a previous child variable in the Cox frailty model which returned the highest coefficient of determination and lowest log-likelihood. The study found parental competence, explains the unseen family effect and other significant programmable predictors were to be considered for a good survival program. It also recommended the use of frailty models when observations are correlated, [12].

Rahman M. and IsIam R. in 2010 discussed and compared various covariates of infant and under-five mortality in the context of overall country, urban and rural levels of Bangladesh with the help of discriminant analysis. The data was obtained

from Bangladesh Demographic and Health Survey done in 2004. The discriminant analysis used the stepwise procedure to rank the only significant variables as per the Wilks lambda values. They also calculated the canonical discriminant function coefficients for the predictor variables thereafter a comparison was done. Based on the analysis, breastfeeding was the most significant variable in discriminating those mothers experiencing infant mortality or not from those experiencing under-five mortality or not. It was also found that discriminant function was statistically significant and discriminates well. In summary, the study suggested that improvements in the health system are essential for promoting the breastfeeding practices (both inclusive and exclusive), seen as effective methods to reach families and communities with targeted messages and information, [15].

Kyalo in 2009 studied the factors influencing child mortality in Somalia with a specific focus on the effect of household's environmental, socio-economic and bio-demographic characteristics on child mortality in Somalia. The variables of the study were the households' wealth status, type and region of residence, mother's level of education, mother age, source of drinking water, type of toilet facility children ever born, current marital status of the mother and sex of the child.

The study used data from the Somali 2006 Multiple Indicator Cluster Survey (MICS) and used descriptive statistics and Cox regression to analyze the data. Using Cox proportional hazard regression, it was found that household's socio-economic, environmental characteristics and bio-demographic variables have significant impact on child mortality in Somalia. Further, it was found that there was lower risk of death experienced among children born in Somali-land as compared to those born from other regions with the risk being higher among boys and children born to mothers then not in marriage unions. The study recommended the use of policies aimed at achieving reduced child mortality in Somalia should be directed at improving the household's socio-economic and environmental status. Further, the government was to prioritize efforts to strengthen national reconciliation and

comprehensive sustainable peace [11].

2.3 Conceptual Framework

Under-five mortality is determined by a number of factors; socio-economic, environmental and bio-demographic. All this determinants of child mortality necessarily operate through a common set of biological mechanisms or proximate determinants that directly influence that risk of mortality. The socio-economic factors further operate through maternal, biological, environmental, nutritional and health seeking behavior factors.

This framework assumes that on a normal setting; over ninety-seven per cent of newborn infants are expected to survive through the first five years of life and that reduction in this survival probability in any given society comes as a result of the operation of social-economic, biological and environmental forces. Growth faltering and ultimately under-five mortality in children is seen as the cumulative consequences of multiple predictor variables originating from the fore-mentioned broad categories.

```
┌─────────────────────────────────────┐
│   DETERMINANTS OF UNDER-FIVE         │
│           MORTALITY                  │
└─────────────────────────────────────┘
```

```
┌──────────────────┐  ┌──────────────────┐  ┌──────────────────┐
│  SOCIO-ECONOMIC  │  │ BIO-DEMOGRAPHIC  │  │  ENVIRONMENTAL   │
└──────────────────┘  └──────────────────┘  └──────────────────┘
```

```
┌──────────────────┐  ┌──────────────────┐  ┌──────────────────┐
│                  │  │ Age of the mother│  │ Source of drinking│
│ Mothers education│  │Current marital   │  │      water       │
│   Wealth level   │  │    status        │  │  Toilet facility │
│ Place of residence│ │ Sex of the child │  │    Region of     │
│                  │  │Number of children│  │    residence     │
│                  │  │    ever born     │  │                  │
└──────────────────┘  └──────────────────┘  └──────────────────┘
```

```
              ┌──────────────────────────┐
              │   UNDER-FIVE MORTALITY    │
              └──────────────────────────┘
```
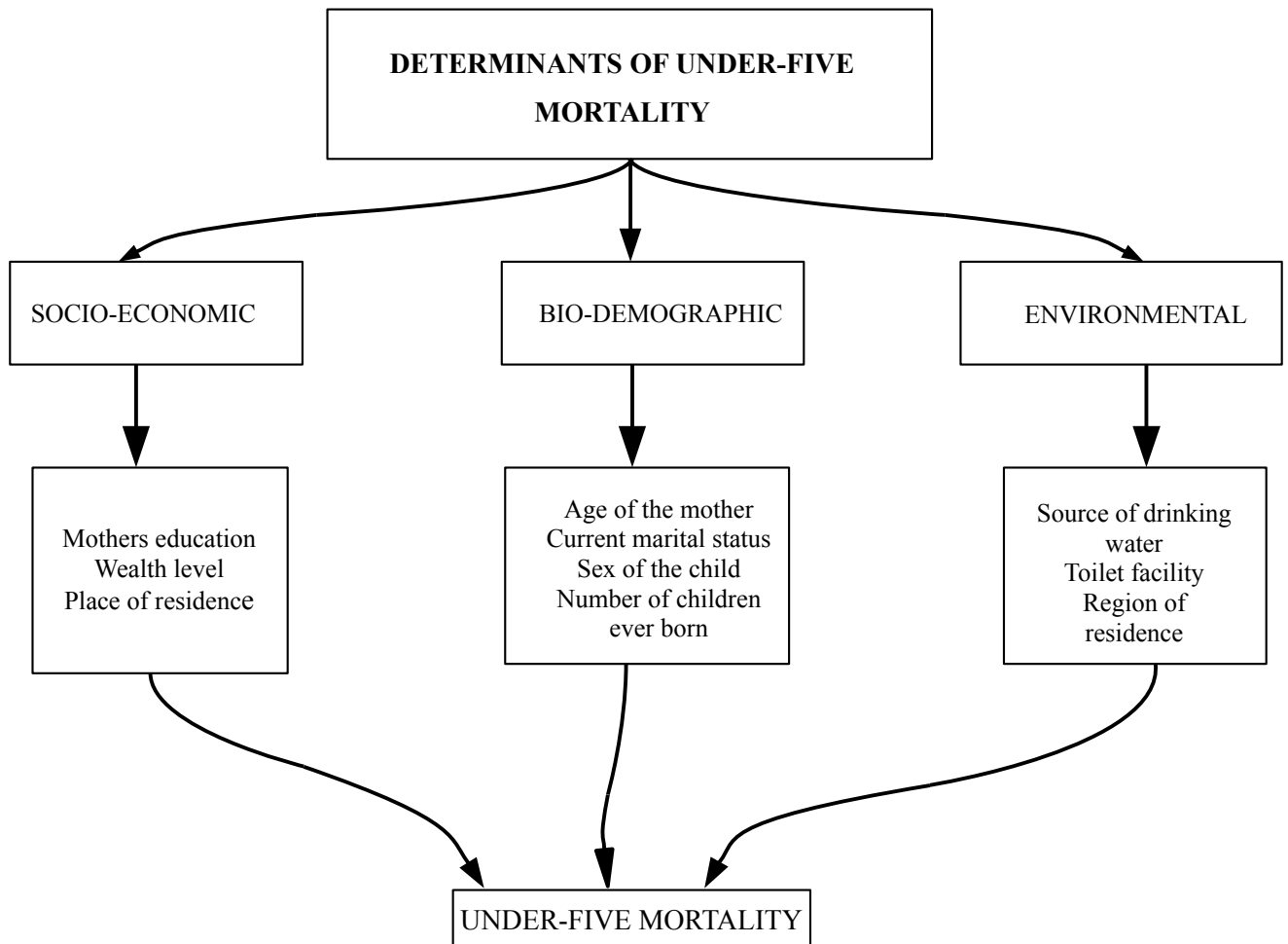
Figure 2.1: Conceptual framework: Determinants of Under-five Mortality

2.4   Summary

In view of the above study, it is evident that there are quite a number of factors that affect the under-five mortality. Most studies reveal that mother's age at birth, place of delivery, sex of the baby and breastfeeding were among the most significant determinants of under-five mortality. Further insight from discriminant analysis Isolated breast-feeding as the most significant predictor. However, these works do differ in terms of recommendations and very few are directly related to Somalia, thus, there is need for more exploration into the determinants of under-five-mortality in Somalia to arrive at a more customized solution for the country.

CHAPTER 3

METHODOLOGY

3.1 Data source

The data for analysis was obtained from UNICEF Somalia. The sample for the
Somali Multiple Indicator Cluster Survey and Pan Arab Project for Family Health
(MICS/PAPFAM) was designed to provide estimates on a large number of indi-
cators on the situation of children and women at the national level, for urban
and rural areas, and for the three zones: North West Zone, North East Zone and
Central South Zone. Zones were identified as the main sampling domains and the
sample was selected in four stages [20].

The target sample size suitable to have unbiased results was determined to be
6,000. The population was then subdivided into three zones which were in turn
divided into clusters.The three zones were the North West, the North East and
the Central south zones. The first two zones had 60 clusters while the latter had
130 clusters. In all the clusters, there were 24 households, [20].

The clusters were allocated to the districts based on the following criteria. The

three zones have regions which intern have districts. The districts were selected using probability proportional to size (PPS) approach. Out of 114 districts in Somalia, 57 were selected. The quantity of clusters allocated to each district was determined based on the estimated population of a particular district. The proportion of urban to rural clusters was determined according to the estimated populations falling within each category within each district. The rural population includes that population which is settled and that which are nomads, [20].

A random selection using probability proportional to size sampling was also used to select households among the permanent and temporary settlements. The temporary settlements were usually taken near water points where nomads water their animals. The sole performance was to target them and include them in the sample size, [20].

Finally, sampling was done to select clusters within the settlements. In cases where the settlements had an estimate of over 150 households, subdivisions was done where into roughly equal sizes of estimated households. Having the subdivisions, a random samples were then done to get the required number of clusters having the required number of households, [20].

3.1.1 Data analysis technique and software used

The variables were recorded using SPSS 21 version. Categorical variables were converted to dichotomous.

The researcher has availed a method for using dichotomous variables known as **dummy variables** which act as replacement variables for the non-metric variable. A dummy variable is a dichotomous variable that represents one category of a non metric independent variable, [9].

3.1.2 Validation

The sample size were divided into two;

(i). Sample for analysis

This can be used to estimate the discriminant function or scores. The analysis sample size is 4,636 which is equivalent to 65% of the total sample size.

(ii). Holdout sample

This can be used to validate the result and to get or to find the overall predictive accuracy of the model called the hit ratio. The holdout sample size is 2,464 which is equivalent to 35% of the total sample size.

3.2 Variable definition

Table 3.1: Definition of the variables

Dependent Variable	Definition
Under-five Mortality	Alive = 1, Dead = 0
Explanatory Variables	
Mothers Education	1 = education, 0 = no education
Place of Residence	1 = Urban, 0 = Rural
Zone	1 = North,0 = South
Wealth class	1 = rich, 0 = poor
Kind of Toilet Facility	1 = has toilet facility, 0 = No toilet facility
Source of drinking water	1 = piped water, 0 = surface water
Children ever born	Continuous variable
Mothers age	Continuous variable Age (15-49)years
Current marital Status	1 = yes currently married, 0 = not in marital status
Sex of the Child	1 = Male, 0 = Female

3.3 Model Formulation and Assumptions

3.3.1 The Objectives of Discriminant Analysis

According to Rencher (2002), there are two major objectives of the discriminant analysis

(i). Linear functions of the variables are used to describe the differences between

two or more groups the relative contribution of the p-variables to the separation of the groups is calculated and an optimal plane found on which the points are projected to best illustrate the group configuration.

(ii). Classification functions which allocate observations to groups (prediction) are derived, that is, an individual sampling unit is allocated to one of the identified groups by means of the classification functions.

3.3.2   Fishers Linear Discriminant Analysis Model

Linear Discriminant Analysis is a classification method originally developed in 1936 by R.A. Fisher. It is based upon the concept of searching for a linear combination of the variables that best separates two classes (groups).

$$Z = V_1 X_1 + V_2 X_2 + ... + V_i X_i \qquad \textbf{Discriminant model} \qquad (3.1)$$

Where

Z = discriminant function

V = discriminant coefficient

X = independent variable

i = Number of the independent variables, [1].

$$J(V) = \frac{V^T \mu_0 - V^T \mu_1}{V^T S_i} \qquad \textbf{Score function} \qquad (3.2)$$

$$V = S^{-1}(\mu_0 - \mu_1) \qquad \textbf{Model Coefficients} \qquad (3.3)$$

where S are the variance-covariance matrices

μ_0 and μ_1 are the mean vectors, [17].

3.3.3 Assumptions of Discriminant Analysis

Fishers approach does not assume that the populations are normal. It does, however, implicitly assume that the population covariance matrices are equal, because a pooled estimate of the common covariance matrix is used, [10].

Testing the assumptions of equal variance-covariance matrices

The null hypothesis is when the variance-covariance matrices of the two groups are the same in the population. For this assumption to hold, the log determinants should be equal. Using BOX's M test, we are looking for a non-significant, [1].

If this assumption of equal variance-covariance matrices is violated, the solution remains to consider a large sample size; then the violation becomes a simple problem and the validity of estimating the discriminant function can be checked by hit ratio of the holdout sample, [18].

No Multicollinearity

The multicollinearity problem can be solved using stepwise discriminant analysis, [18].

The sample size of the smallest group needs to exceed the number of predictor variables. As a rule of thumb, the smallest sample size should be at least 20 for a few (4 or 5) predictors, [14].

3.3.4 Partitioning Sums of squares (SS) in Discriminant Analysis

In Discriminant analysis

The Total SS $\sum (Z_i - \overline{Z})^2$ is partitioned into:

Within Group SS $\sum (\overline{Z}_i - \overline{Z})^2$

Between Group SS $\sum(Z_ij - \overline{Z_j})^2$

$$\sum(Z_i - \overline{Z})^2 = \sum(\overline{Z}_i - \overline{Z})^2 + \sum(Z_ij - \overline{Z_j})^2$$

$i =$ an individual case, $j =$ group j

$Z_i =$ individual discriminant score

$\overline{Z} =$ overall mean of the discriminant scores

$\overline{Z}_j =$ mean discriminant score for group j, [6].

3.3.5 Stepwise Discriminant Analysis

The method for entering the predictor variables into the model is the stepwise discriminant analysis. Stepwise discriminant analysis picks only the most significant variables.

The importance of the variables in discriminant analysis can be studied using the method suggested by Huberty (2006). The most important variable is the one which has the largest Wilk's lambda value. The second most important variable is the one which has the second largest Wilk's Lambda value. So the variables are ranked according to their importance depending on the ranks of the lambda values.

Wilk's lambda is the ratio of within sum squares to the total sum squares.

$$\frac{\text{within sum squares}}{\text{Total sum squares}} = \frac{\sum(z_{ij} - \overline{z_j})^2}{\sum(z_i - \overline{z})^2} \tag{3.4}$$

Where

$i =$ an individual case

$z_i =$ individual discriminant score

$\overline{z} =$ overall mean of the discriminant scores

$\bar{z}_j$ = mean discriminant score for group j

3.4 Specification of the models

$$Z = V_1 X_1 + V_2 X_2 + \ + V_i X_i \tag{3.5}$$

Where

Z =under-five mortality

V_i =weight or discriminant coefficient

X_1 = age of the mother

X_2 = children ever born

$X_3 = $ current marital status of the mother

$X_4 = $ source of drinking water

$X_5 = $ Region of residence

$X_6 = $ toilet facility

$X_7 = $ Place of residence

$X_8 = $ mother's education

$X_9 = $ wealth level

$X_{10} = $ sex of the child

Discriminant Function Analysis is used when the dependent variable is categorical for example child mortality is binary dependent variable predicting whether the child will die or not as a function of various socio-economic, bio-demographic and environmental factors.

3.5 Estimation of the parameters of the model

We are using Fisher's linear Discriminant to estimate the parameters. Fisher suggested a method which makes the mean of the two groups as far apart as possible and the variance as closer as possible.

$$Z = V_1 X_1 + V_2 X_2 + ... + V_i X_i = V^T X \tag{3.6}$$

The idea is to find the direction V such that the data of the two groups are well separated if the projected onto the direction.

Consider we have two groups.

Let $\{(X_i, Z_i), i = 1, ..., n\}$ be the data.

Let $Z_i \in \{0, 1\}, G_0$ and G_1 denote the two groups. Thus, if $Z_i = 0$ then $X_i \in G_0$ and if $Z_i = 1$ then $X_i \in G_1$. Also n_0 and n_1 denote the number of samples of each group $(n = n_0 + n_1)$.

Z_i are one dimensional data we get after the projection .

Let μ_0 and μ_1 be the means of the data from the two groups

$$\mu_0 = \frac{1}{n_0} \sum_{X \in G_0} X_i, \mu_1 = \frac{1}{n_1} \sum_{X \in G_1} X_i \tag{3.7}$$

The corresponding means of the projected data would be

$$\mu_0 = V^T \mu_0 \text{ and } \mu_1 = V^T \mu_1 \tag{3.8}$$

The difference $(\mu_1 - \mu_0)$ gives us an idea of the separation between samples of the two groups after projecting the data onto the direction V.

We may want V that maximizes $(\mu_1 - \mu_0)^2$ and also minimizes the within group variances $(s_0^2 + s_1^2)$.

Define

$$s_0^2 = \sum_{X_i \in G_0} (V^T X_i - \mu_0)^2, s_1^2 = \sum_{X_i \in G_1} (V^T X_i - \mu_1)^2 \tag{3.9}$$

These gives us the variances of the two groups in the projected data. We want large separation between μ_0 and μ_1 relative to the variances.

Hence we take our objective to maximize

$$J(V) = \frac{(\mu_1 - \mu_0)^2}{s_0^2 + s_1^2} \tag{3.10}$$

We can write J into a more convenient form. We have

$$(\mu_1 - \mu_0)^2 = (V^T \mu_1 - V^T \mu_0)^2$$
$$= V^T(\mu_1 - \mu_0)(\mu_1 - \mu_0)^T V$$

$$(\mu_1 - \mu_0)^2 = V^T(\mu_1 - \mu_0)(\mu_1 - \mu_0)^T V$$

Thus we have $(\mu_1 - \mu_0)^2 = V^T S_B V$

Where

$$S_B = (\mu_1 - \mu_0)(\mu_1 - \mu_0)^T$$

S_B is called between scatter matrix.

We have

$$s_0^2 = \sum_{X_i \in G_0} (V^T X_i - V^T \mu_0)^2$$

$$= \sum_{X_i \in G_0} V^T (X_i - \mu_0)^2$$

$$= \sum_{X_i \in G_0} V^T (X_i - \mu_0)(X_i - \mu_0)^T V$$

$$s_0^2 = \sum_{X_i \in G_0} V^T (X_i - \mu_0)(X_i - \mu_0)^T V \qquad (3.11)$$

Similarly, we can get

$$s_1^2 = \sum_{X_i \in G_1} V^T (X_i - \mu_1)(X_i - \mu_1)^T V$$

Thus we can write

$$s_0^2 + s_1^2 = V^T S_W V \qquad (3.12)$$

Where

$$S_W = \sum_{X_i \in G_0} V^T (X_i - \mu_0)(X_i - \mu_0)^T V + \sum_{X_i \in G_1} V^T (X_i - \mu_1)(X_i - \mu_1)^T V$$

S_W is called within group scatter matrix.

Hence we can write J as a function of V

$$J(V) = \frac{(\mu_1 - \mu_0)^2}{s_0^2 + s_1^2} = \frac{V^T S_B V}{V^T S_W V} \qquad (3.13)$$

We want to find a V that maximizes $J(V)$.

$$J(V) = \frac{V^T S_B V}{V^T S_W V}$$

Differentiating with respect to V and equating to zero

$$\frac{d}{dV}J(V) = d/dV(\frac{V^T S_B V}{V^T S_W V}) = 0$$

$$(V^T S_W V)(\frac{d(V^T S_B V)}{dV} - (V^T S_B V)\frac{d(V^T S_W V)}{dV} = 0$$

$(V^T S_W V)2S_B V - (V^T S_B V)2S_W V = 0$

Dividing by$(V^T S_W V)$

$(\frac{V^T S_W V}{V^T S_W V})S_B V - (\frac{V^T S_B V}{V^T S_W V})S_W V = 0$

$S_B V - J S_W V = 0$

$S_W^{-1} S_B V - JV = 0$

Solving the generalized eigenvalue problem

$$V = S_W^{-1}(\mu_1 - \mu_0) \tag{3.14}$$

3.6 The significance of the discriminant model

To test the significance of the discriminant function we use some equations by Nasir Uddin et al. (2013).

The eigenvalue tells us the discriminatory power of the model or the ability of the discriminant function. Eigenvalue is the ratio between sum squares to within sum squares.

$$\text{Eigenvalue } = \frac{BSS}{WSS} = \frac{\sum (\bar{z}_j - \bar{z})^2}{\sum (z_{ij} - \bar{z}_j)^2} \tag{3.15}$$

The larger the value of lambda, the greater the discriminatory power of the model. From the description of the eigenvalue above, we can deduce two tests, thus the canonical correlation and the Wilk's lambda.

(1). Canonical Correlation (η), which is the correlation of predictors with the discriminant scores in the model is given by;

$$\text{The canonical correlation } (\eta) = \sqrt{\frac{\lambda}{1+\lambda}} = \sqrt{\frac{BSS}{TSS}}$$

(2). Wilk's lambda ($\wedge$), which is given by the expression:

$$\wedge = 1 - \eta^2 = \frac{1}{1+\lambda} = \frac{WSS}{TSS}$$

It can be converted into a chi-square statistic

Chi-square, $\chi^2 = -[(n-1) - 0.5(m + p + 1)]ln\wedge$

$df = p - 1$

n =sample size

m =number discriminant extracted

p =number of predictor variables

In the chi-square test, the null hypothesis (H_0) is that the centroid of the two groups and the overall centroid are all equal. At 1% level of significance, the rejection of the null hypothesis means that the estimating discriminant function and interpretation of the results are statistically significant.

3.7 Classification

Classification matrices are constructed to determine the predictive accuracy of a discriminant function. The classification matrix is a matrix containing numbers that reveal the predictive ability of the discriminant function. The numbers of correct classifications are found on the diagonal of the matrix whereas the off-diagonal numbers represent misclassifications, [9].

3.7.1 Optimal cutting score

The optimal score is used to categorize two groups discriminant function into two groups. It is thus used to construct the classification matrix.

The formula for the cutting score is given by Ramayah [16] in 2010 as:

Equal group

$n_0 = n_1$

$$C = \frac{n_0 \overline{Z}_0 + n_1 \overline{Z}_1)}{(n_0 + n_1)} \qquad (3.16)$$

where

C = optimal cutting score for equal group size

n_0 = number of observations in Group 0

n_1 = number of observations in Group 1

$\overline{Z}_0$ = Centroid for Group 0

$\overline{Z}_1$ = Centroid for Group 1

Unequal group

$n_0 \neq n_1$

$$C = \frac{(\overline{Z}_0 + \overline{Z}_1)}{2} \tag{3.17}$$

where

C = optimal cutting score for unequal group size

$\overline{Z}_0$ = Centroid for Group 0

$\overline{Z}_1$ = Centroid for Group 1

If $Z_i \leq C$ classify the observation i to group 0.

If $Z_i > C$ classify the observation i to group 1.

3.7.2 Statistical significance for prediction

There are three criteria that can be used to test whether the model developed has good predictive accuracy (Ramayah et al., 2010).

1. Maximum Chance Criteria (MCC)

    Predict that all cases are in the group with the largest of cases.

    $$MCC = \frac{n_L}{N_L}$$

    where

    n_L = number of subjects in the larger of the two groups

    N_L = number of subjects in to combined groups

2. Proportional Chance Criteria

    It randomly classifies the cases proportionate to the number of cases in either

group.

$$C_{PRO} = P^2 + (1 - P)^2$$

$1 - P =$ Proportion or individuals in group 1

$P =$ Proportion of individuals in group 0

3. Press Q statistic
$$Q = \frac{(N - (n \times k))^2}{N(K - 1)}$$

where,

$Q \sim \chi^2$ with one degree of freedom

$N =$ Total sample size

$n =$ Number of observations correctly classified

Press Q statistics is used to test whether the model hit ratio is significantly better than chance. Most researchers would accept a hit Ratio that is 25% larger than that due to chance, [1].

DATA ANALYSIS AND RESULTS

4.1   Test of equality of covariance matrices

Table 4.1: Test of equality of covariance matrices by using BOX'S M test

U5M	Rank	Log determinants	BOX'S M	Approx.F	Df1	Df2	Sig.
Alive	5	-1.661	200.690	13.363	15	51419164.661	0.001
Dead	5	-1.120					
Pooled within-groups	5	-1.433					

H_0:the variance-covariance matrices of the two groups are the same

For this assumption to hold, the log determinants should be equal. When tested Box's M, we are looking for a non-significant M to show similarity and lack of significant differences. In this case the log determinants appear similar and Box's M is 200.690 with $F = 13.363$ which is significant at $p < 0.001$ (tables 4.1)

however,with large samples, a significant result is not regarded as too important .

4.2 Wilks Lambda and P-values for the Significant Predictor Variables

Table 4.2 shows the Wilk's Lambda and the corresponding p-values for the 5 variables selected through the stepwise procedure. From the results in Table 4.2, we observe that children ever born is the most significant variable, based on Wilk's lambda, in discriminating for the two groups, whether child is in the alive or dead group. This is because child ever born provides the highest Wilk's lamda value. The second most significant variable is the water source then age of woman, marital status and region of residence in that order.

Table 4.2: Wilks Lambda and P-values for the Significant Predictor Variables

Predictor variables	Wilks lambda values	P-values
Children ever born	0.830	0.001
Source of drinking water	0.815	0.001
Age of mother	0.805	0.001
marital status	0.800	0.001
Region of residence	0.797	0.001

4.3 Significance of the Discriminant Function

The calculated values for determination of the significance of the discriminant function are shown in Table 4.3

Table 4.3: Significance of the discriminant function

Function	Eigen value	% of variance	Cumulative %	Canonical correlation	Wilk's $\wedge$	χ^2	df	Sig.
Under-Five Mortality	0.262	100	100	0.456	0.792	1080.031	5	0.001

Table 4.3 shows that the discriminant function is statistically significant and it discriminates well by explaining about 100% of the variation between the two groups. The eigenvalue 0.262 is related to the discriminating power of the function. This value is an indication that discrimination between the two groups exist. The canonical correlation (0.456) represents a measure of association that summarizes the degree of relatedness between the selected groups. A value of zero indicates no association at all, however, a large number (positive) indicates presence of association up to a maximum value of 1.

4.4 Canonical Discriminant Coefficients

The importance of the set variables that have been picked from the stepwise procedure is observed through the study of standardized canonical discriminant function coefficients as shown in Table 4.4. The magnitude of the coefficients translate to how important the variable is as a discriminant. From the table children ever born has the highest as was earlier indicated by the Wilks Lambda.

Table 4.4: Canonical Discriminant Function Coefficients

Under-five mortality	Standardized coefficients	Unstandardized coefficients
Age of mother	-0.269	-0.053
Children ever born	1.185	0.520
Current marital status	-0.171	-0.659
Resion of Residence	-0.138	-0.281
Water source	-0.343	-0.688
Constant		0.455

Standardized canonical discriminant coefficients give us the importance of each predictor variable. It is similar to Wilks lambda. Using the standardized canonical coefficients, the fitted discriminant model is as follows:

$$Z = 1.185 \text{ children ever born } -0.343 \text{ Source of drinkinking water } -0.269 \text{ age of mother}$$

$$-0.171 \text{ current marital status} - 0.138 \text{ region of residence}$$

Unstandardized canonical discriminant coefficients explain group membership or prediction. Using these coefficients as shown in Table 4.4, the fitted prediction model is as follows:

$$Z = 0.520 \text{ children ever born } - 0.688 \text{ source of water } - 0.053 \text{ age of mother}$$

$$-659 \text{ current marital status } - 0.281 \text{ Region of residence } + 0.455$$

4.5 Group Centroid

The Centroid is the mean value of the discriminant scores for a particular group. The centroid is calculated from the fitted prediction model. The centroid for the alive (did not die) is -0.373 and Dead (died) is 0.707 as shown in Table 4.5.

Table 4.5: Calculated Group Centroid

Under-five mortality	Function 1
Alive	-0.373
Dead	0.707

From the group centroids, a cut off value is calculated in order to come up with a classification rule. Using the values in Table 4.5, the calculated cut off value is:

Cut off value $= \frac{\overline{Z}_0 + \overline{Z}_1}{2} = \frac{0.373 + 0.707}{2} = 0.54$

Cut off value $= 0.54$

The 0.54 calculated is used as threshold for classifying a child to either of the two groups. If z score for a given child is greater than 0.54, then the child would be assigned to group 1 which represents dead, otherwise, the child is assigned to group 0, representing alive. For example, a child from a household whose mother has had 8 children ever born, water source is surface water, age of mother is 32 years, current marital status of mother is married, region of residence is north; the z score would be:

$$Z = 0.520(8) - 0.688(0) - 0.053(32) - 0.659(1) - 0.281(1) + 0.455$$

$$Z = 1.979$$

The calculated Z score (1.979) is compared to the cut off value (0.54) in order to assign the child to either group.

$$1.979 > 0.54$$

Since the calculated z score is greater than the cut off value, we assign the child into group 1 (dead).

4.6 Classification Accuracy of Fitted Prediction Model

Using the fitted prediction model on the hold out sample (2,464), the classification matrix developed is as shown in Table 4.5.

Table 4.6: Classification table Based on the Holdout Sample

Actual Group	No. of casesB	Predicted group Membership	
		Alive	Dead
Alive	1,612	1,434 (89%)	178 (11%)
Dead	852	467 (54.8%)	385 (45.2%)

From the classification matrix, the hit ratio is calculated. The hit ratio gives the overall prediction accuracy of the model. It is calculated as: numbers on the diagonal of the classification matrix divided by the total of the number of cases:

Hit ratio= $\frac{1434+385}{2464} = 0.738 = 73.8\%$

Hit ratio $= 73.8\%$

Therefore, the percentage of grouped cases correctly classified in the hold out sample is 73.8%.

To determine the goodness of fit of the model developed, the calculated values for three criteria are hereby presented:

1. Maximum Chance

$$C_{max} = \text{Size of largest group} = \frac{1612}{2464} = 0.65$$

37

2. Proportional Chance

$$C_{PRO} = P^2 + (1 - P)^2$$

where p is the proportion of individuals in group 1

$$(0.6)^2 + (1 - 0.65)^2 = 0.55$$

3. Press Q statistic

$$Q = \frac{(N - (n \times k))^2}{N(K - 1)}$$

where

$Q \sim \chi^2$ with one degree of freedom

$N = $ Total sample size

$n = $ Number of observations correctly classified

$$\text{Press } Q = \frac{(2464 - (1818 \times 2))^2}{2464(2 - 1)} = 557.46$$

The calculated three criteria are compared to the hit ratio as shown in Table 4.6

Table 4.7: Comparison of the three Goodness of fit Criteria

Measure	value	Hit ratio
Maximum chance	0.65	73.8
Proportional chance	0.55	73.8
Press Q table value	6.635	
Press Q calculated value	557.46**	

**$P < 0.01$

From the results above, and insights from the appendix III, the predictive accuracy of the model; for the analysis sample was 73.3% while that of the holdout sample was 73.8%. The values in Table 4.6 indicate that the hit ratio of 73.8% for the holdout sample exceeded both the maximum and proportional chance values. The Press Q statistics of 557.46 was significant. Thus, the model constructed has a good predictive power.

4.7   Group Statistics

Group means and standard deviations for each variable for Alive and Dead group are calculated in Appendix I. Group mean provides an idea about whether the means of the variables differ between the groups. In addition, group means and group standard deviations can be used as characteristics profile for the two groups.

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Discriminant analysis was used to identity significant factors contributing to U5M in Somalia. Using the stepwise discriminant analysis procedure, children ever born, source of drinking water, age of the mother, current marital status of the mother and region of residence were found to be significantly contributing to U5M. The discriminant function was statistically significant based on fit statistics used which included eigenvalue, canonical correlation and the transformed chi-square statistic. A predictive model was fitted using these variables and based on hold out sample of 2,464 observations, 73.8% of the cases were correctly classified. Therefore, the discriminant function and the predictive model constructed were adequate and thus can be used to classify a child into any of the two groups based on the significant factors identified.

5.2   Recommendation

From the results of identified significant factors, the following measures if addressed by the government and other stakeholders would help reduce U5M in Somalia :

(i). enhance education facilities and curriculum for the mothers and girl child education in general;

(ii). provision of safe drinking water for the households in general;

(iii). strengthen maternal and child health care especially at community level in all the regions of Somalia

Last but least, further research could be conducted to model health related factors contributing to U5M in Somalia.

Bibliography

[1] Agrest A. (1996). *An Introduction to categorical data analysis* John Wiley and Sons Inc.

[2] Alvin C. Rencher (2002). *Methods of multivariate analysis.* 2nd ed. John Wiley and Sons Inc.

[3] Boachie-Yiadom S, Kwaku F. D. & Tawiah R. (2014). Analysis of Under-five Mortality in Ghana Using Logit Model, *International Journal of Statistics and Applications* Vol. 4, No.4: 192-197

[4] Chowdhury A. H. (2014). Determinants of Under-Five Mortality in Bangladesh. *Open Journal of Statistics.* Vol. 3 pp.213-219.

[5] ECA. (2013) *Report on Progress in achieving the Millennium Development Goals in Africa,2013.*

[6] Friel M. C. l (n.d). *Notes on Discriminant Analysis.* Criminal Justice Center, Sam Houston State University.

[7] Geneti K.T. & Deressa T.K. (2014). Correlates of under-five mortality in Ethiopia. A logistic Regression Analysis *Avances and applications in Statistics.* Vol. 39, No.1, pp. 61-79. Pushba Publishing House,Allahabad,India

[8] Huberty C. J. (2006). *Applied MANOVA and discriminant analysis*, 2nd ed. New York; John Wiley and Sons Inc.

[9] Hair J.F. et al (2009). Multivariate Data Analysis, 7th ed.

[10] Johnson R. A. & Wichern D.W. (2007) *Applied Multivariate Statistical Analysis*

[11] Kyalo N. C. (2009).*Factors incluencing child mortality in Somalia.* Thesis, University of Nairobi.

[12] Mani K., Dwivedi N.S. & Pandey M.R. (2012). Determinants of Under-Five Mortality in Rural Empowered Action Group States in India: An Application of Cox Frailty Model. *International Journal of MCH and AIDS (2012).* Vol. 1, No. 1, pp. 60-72

[13] MDGs. (2014). *Millennium Development Goals, Report for the Levels and Trends in child mortality, 2014*

[14] Rahman A. A. H. et al (2009). Discriminant Analysis of Animal Species Odors Response *Advances in Natural Science* . Vol. 2, No. 2

[15] Rahman M. & IsIam R. (2010). Infant and Under-five Mortality in Bangladesh: Discriminant Analysis. *Chinese Journal of Population Resources and Environment.* Vol.8, No.4 pp.79-84

[16] Ramayah et al, 2010, Discriminant Analysis; An illustrated example: *African Journal of Business management.* Vol. 4, No.9, pp. 1654 - 1667.

[17] Saed Sayad. (2010). *Linear Discriminant Analysis (LDA).* University of Toronto viewed on 5th June 2015. Available at <www.Saedsayad.com>.

[18] Uddin Nasir et al (2013). Discriminant Analysis as an aid to human resource selection and human resource turnover minimization decisions. *International journal of Business and management.* Vol. 8, No. 17.

[19] UNDP (2012). *Millennium Development Goals, Reduce child mortality in Somalia.*

[20] UNICEF. (2006). *Monitoring the situation of Children and Women*: Multiple Indicator Cluster Survey Report in Somali

[21] United Nations (2007) *World population Prospects: The 2006 revision,* CD-ROM Edition. United Nation's Children's Fund(2005). *The state of the world's Children.* UNICEF, New York.

[22] WHO (2012). *Child health in Somalia: Situation Analysis.*

APPENDIX I

Analysis Case processing Summary

Table 5.1: Sample for analysis and holdout

Unweighted Cases		N	Percent
Valid		4636	65.3
	Missing or out-of-range group codes	0	0
	At least one missing discriminating variable	0	0
Excluded	Both missing or out-of-range group codes and at least one missing discriminating variable	0	0
	Unselected	2464	34.7
	Total	2464	34.7
Total		7100	100.0

APPENDIX II

Table 5.2: Group statistics

Under five mortality		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
Alive	Age of woman	27.65	6.403	2990	2990.000
	Children ever born	3.77	2.144	2990	2990.000
	Marital status	0.93	0.250	2990	2990.000
	Resion of Residence	0.46	0.499	2990	2990.000
	Water source	0.84	0.367	2990	2990.000
	Age of woman	30.50	6.513	1646	1646.000
	Children ever born	5.91	2.494	1646	1646.000
Dead	Marital status	0.93	0.252	1646	1646.000
	Resion of Residence	0.33	0.472	1646	1646.000
	Water source	0.75	0.435	1646	1646.000
	Age of woman	28.66	6.584	4636	4636.000
	Children ever born	4.53	2.493	4636	4636.000
Total	Marital status	0.93	0.251	4636	4636.000
	Resion of Residence	0.42	0.493	4636	4636.000
	Water source	0.81	0.395	4636	4636.000

APPENDIX III

Table 5.3: Classification table

			Under five mortality	Predicted Group Membership		Total
				Alive	Dead	
Cases Selected	Original	Count	Alive	2661	329	2990
			Dead	910	736	1646
		percentage	Alive	89.0	11.0	100.0
			Dead	55.3	44.7	100.0
Cases not selected	Original	Count	Alive	1434	178	1612
			Dead	467	385	852
		percentage	Alive	89.0	11.0	100.0
			Dead	54.8	45.2	100.0

(a). 73.3% of selected original grouped cases correctly classified (from the analysis sample).

(b). 73.8% of unselected original grouped cases correctly classified (from the hold-out sample).