KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

IDENTIFYING PREDICTIVE LATENT VARIABLES FOR UNDER-5 MALNUTRITION IN GHANA.

By

Kofi Asare-Bawuah (BSc. Actuarial Science)

(PG6200911)

A Thesis submitted to the Department of Mathematics, Kwame Nkrumah University of Science and Technology in partial fulfillment of the requirements for the degree of

Master Of Philosophy

College of science

MAY 2013

Declaration

I hereby declare that this submission is my own work towards the award of the M.Phil degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.

Kofi Asare-Bawuah	....	01/10/2013
Student	Signature	Date

Certified by:		
Dr. Atinuke O. Adebanji	....	02/10/13
Supervisor	Signature	Date

Certified by:		
Prof. S.K. Amponsah	....	21/10/13
Head of Department	Signature	Date

Acknowledgement

Thank You to the Almighty God for taking me through the two year programme successfully. To Dr. Atinuke O. Adebanji for her time, guidance and patience throughout the research. To the Morny family, for their hospitality during my two-year stay with them for this programme. To my programme colleagues Odame Owiredu Emmanuel and Magdalene Torto for their immense assistance during the two year programme.

Dedication

To Almighty God , my late father WG. CDR. (RTD) Joseph Bawuah, my dear mother, Madam Cecilia Asantewah and my lovely sister Evelyn Bawuah.

Abstract

This study assessed the latent factors associated with Malnutrition among children under five in Ghana. Malnutrition has been an age old problem to most developing countries, and in Ghana, has been a detrimental factor to the survival of children especially in their early years when they are most susceptible to many health problems. It has been estimated that approximately 54% of all deaths among under-five children were associated with PEM, making this the single utmost cause of child mortality in Ghana. The 2008 GDHS data on children was used for this study, from which 22 variables were employed for the analysis. The Bartlett's Test of Sphericity (a measures of singularity of the matrix of correlations among these 22 variables) gave a p-value=0.000 indicating that the correlation matrix used for this analysis was not an identity matrix and that the variables had significant association among themselves. The Kaiser-Meyer-Olkin measure of sampling adequacy gave a value of 0.707 indicating the appropriateness of the sample for Factor Analysis. Four Latent Factors were obtained using the Principal Component method of factor extraction. These Latent Factors are 'Child growth per Breastfeeding', 'Wealth', 'Vitamin A level', and 'Social factors'. It can be inferred from the findings that interventions on malnutrition should address these Latent Factors individually and not in a congregate manner.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter consists of a brief background of the study, an introduction to the concept of malnutrition, Governmental Policies aimed at reducing the incidence of malnutrition, and profile of Ghana. It will also take a look at the problem statement of malnutrition in Ghana, research objectives, justification, scope and limitations of the study as well as organization of the thesis.

The long-term vision for Ghana is to become a middle-income country by year 2020. Though year 2020 is not too distant from today, the realization of this national policy can only be apparent when children being born today are given the opportunity to live and attain their full potential. This study assessed the latent factors associated in determining Malnutrition status among children under 5 years in Ghana. Data from the Ghana Demographic and Health Survey (GDHS) conducted in 2008 was used; where the various determinants of the nutritional status of children under-five were assessed. Factors included: anthropometric measurements of children under 5 years, occupation of the parents, marital status, family income, parental education, maternal nutritional knowledge, residence location (urban or rural), gender, and breastfeeding practices.

Malnutrition has been an age old problem for most developing countries, and in Ghana, has been a detrimental factor to the survival of children especially in their early years when they most susceptible to many health problems. The incidence of malnutrition in Ghana has been most dominant in the form of Protein Energy Malnutrition (PEM), which causes growth impedance and underweight. It has been estimated that approximately 54% of all deaths among under-five children were associated with PEM, making this the single utmost cause of child mortality in Ghana. (Ghana Health Service, 2005).

1.1 Background of the Study

A child who has access to adequate food supply, proper health and care is usually well-nourished. The anthropometric measurements of such a child will be normal, that is the weight and height measurements of that child compares very well with the standard normal distribution of heights (H) and weights (W) of healthy children of the same age and sex. Therefore, the most appropriate method in evaluating the nutritional status and general health of a child is to compare the child's growth indices with the approximate cut-off points in the standard normal distribution of well nourished children that are associated with adequate growth, (Hien and Hoa, 2009)

Malnutrition and its attendant deficiencies with nutrition affect children under-5 years more than any other age group. Since infancy is the most vulnerable period in the stages of life, nutritional insufficiencies will result in the debilitating of the growth of the body, (Shi and Singh, 2009). If this nutritional insufficiency is continued for a protracted period of time, it results in the growth irresolute which is exhibited in the form of low weight, small height, low IQ, (Bhavsar et al. 2012). The future of every country is determined by its growing generation, therefore the health status of children of any country that represents the health status of people of that country. Since this growing generation is going to be the future productive citizens, they should be healthy enough to make use of the full potential of their productive age. There has been much scientific studies in child-health which has provided evidence that the effects of chronic malnutrition beyond the age of 2-3 years, are irreversible, (MotherChildNutrition, 2013).

According to the Food and Agriculture Organization (FAO) report in 2011, The State of Food Insecurity in the World, 2 billion people in the world suffer from various forms of malnutrition. About 2.6 million children die each year as a result of malnutrition which represents a third of child deaths globally. Child malnutrition is one of the most prevalent causes of infant and child mortality. In Sub-Saharan Africa, it accounts for about 2 percent of deaths and about 3 percent of disability-adjusted life years (DALYs) in under-five children. (Nemer L, et al. 2001)

Food security exists when, at all times, everyone has access to and control over sufficient quantities and quality of food needed for an active and healthy life. A measure of household well-being status and determining factor of child survival is the nutritional status of under-five children. For a household this means the ability to secure adequate food to meet the dietary requirements of all its members, either through their own production or through food purchases. Factors affecting food purchases include household income and assets as well as food availability and price in local markets. Food production depends on a wide range of factors, including access to fertile land, availability of labour, appropriate seeds and tools and climatic conditions, (WFP and CDC 2005).

Most major food ~~and nutrition~~ crises do not occur because of a lack of food, but rather because people are too poor to obtain enough food. In Sub-Saharan Africa extreme poverty, which went up from 41 percent in 1981 to 46 percent in 2001 (WorldBank 2004), combined with growing population increased the number of people living in extreme poverty from 231 million to 318 million, making access to food very difficult, thus compounding the problem of malnutrition. (The Independent, London 2007).

The main nutrition problems include inadequate intakes of energy and protein, iodine deficiency disorders, iron deficiency anaemia and vitamin A deficiency. Although these problems are enormous, their full magnitude is unappreciated because usually there are no obvious signs of the problems, and the victims themselves are not aware. As a result not enough attention is paid to malnutrition. Adequate nutrition requires three complementary inputs, that is, caring practices, (such as exclusive breastfeeding and appropriate complementary feeding in infancy), the protection of child health and the provision of adequate household food security. All are necessary to ensure good nutrition. (Oocities, 2013)

1.2 Concept of Malnutrition and its Causes

The term Malnutrition is commonly used to refer to any state in which the body does not receive adequate nutrients for right function (MedicineNet, 2013). Therefore in more modern science of nutrition, a focal construct has been the definition and concept of malnutrition as a state or condition which is brought about by the insufficient intake of nutrients to meet the biological and functional requirements of the human body. So the condition in which the body does not get the right amount of the vitamins, minerals, and other nutrients it needs to maintain healthy tissues and organ function is also termed as malnutrition.

Technically the prefix 'mal' actually refers to both over- and under-intake. However the typical usage until recently had been directed to understanding inadequate intakes of macro- and micronutrients. Malnutrition then becomes an imbalance or deficiency of nutrients. This can come from not having balanced diets or by using up too many nutrients through activities. Malnutrition can be identified by using body weight, body fat, protein stores and laboratory values, (MedicineOnline, 2013). Of central concerns are measurement to which malnutrition is assessed; observation through analysis of biological tissues (e.g., serum), observation of well-established physical attributes (e.g., anthropometric) and clinically observable consequences (e.g., blindness, stunted growth), and by inference from data on admission to health facilities. For example, anthropometric status is commonly used to assess malnutrition of children under age 5 (de Onis et al. 2004). Macronutrient deficiencies occur when there is reduction in macronutrient intake and a concurring decrease in activity and an increased use of reserves of energy (muscle and fat), or decreased growth. Consequently, malnutrition acquired a central role in scientific conceptualization of nutrition; it was often mentioned jointly with the idea of hunger, to the point at which the two often became virtually synonymous. (Wunderlich and Norwood, 2006).

Malnutrition may also occur because a person cannot properly digest or absorb nutrients from the food they consume, as may occur with certain medical conditions. (UNICEF, 2012). All humans require a certain number of calories maintain metabolic functions and fuel activity requirements. However, the minimum number of calories required varies by age, sex, and individual. (Wiley and Allen, 2009). The situation when there is inadequate caloric intake may also result in malnutrition. The impacts of malnutrition on the body vary from mild, or even latent, in which an individual may not experience any physical symptoms of being malnourished . Severe malnutrition may result in permanent, irreversible damage to the body. For instance, a severe iodine deficiency may result in irreversible brain damage even ~~after an~~ individual receives treatment. Starvation is another severe form of malnutrition that occurs when a person doesn't consume enough food to fuel their body. (Wikispaces, 2013). According to a UNICEF(2010) report, a number of different nutrition disorders may arise, depending on which nutrients are under, or over abundant in a person's diet. In most of the world, malnutrition is present in the form of under nutrition, which is caused by a diet lacking adequate calories and protein. While malnutrition is more prevalent in developing countries, it is also present in industrialized countries. In wealthier nations it is more likely to be caused by unhealthy diets with excess energy, fats, and refined carbohydrates. A growing trend of obesity is now a major public health concern in lower socio-economic levels and in developing countries as well. (UNICEF, 2012)

Malnourished children who survive tend to start school late, are more likely to drop out, and have lower adult earnings because of the adverse effect this condition has on the IQ. The ripple consequence is that of compromised human capital, robbing many developing countries of at least 2-3% of economic growth. Investments targeted between birth until two years of the child's life are most desirable because they target the most vulnerable, and prevent irreparable damage to human capital. (WorldBank, 2013)

There are three commonly used measures for detecting malnutrition in children: stunting (extremely low height for age), underweight (extremely low weight for age), and wasting (extremely low weight for height). These measures of malnutrition are interrelated, but studies for the World Bank found that only 9 percent of children exhibit stunting, underweight, and wasting. (Wagstaff and Watanabe, 1999). According to a 2008 review an estimated 178 million children under age 5 are stunted (most of who live in sub-Saharan Africa), 55 million children are wasted, including 19 million who have severe wasting or acute malnutrition, (Bhutta et al, 2008). Measurements of a child's growth provide the key information for the presence of malnutrition, but weight and height measurements alone can lead to failure to recognize kwashiorkor and an underestimation of the severity of malnutrition in children. (Duggan et al, 2008)

A paradigm shift in Ghanaian health policy has been taking place since 2006.

One of the fundamental hypotheses of this the new health policy in Ghana with its theme "Creating Wealth through Health", was that improving health and nutritional status of the population would lead to improved productivity, economic development and wealth creation. (GHS, 2007). This policy adopts an approach which addresses the broader determinants of health. It has generated interest in socio-economic inequalities in health and malnutrition. It was further recognized that not paying attention to malnutrition inequalities during the early years of life is likely to perpetuate inequality and ill health in future generations and thus defeat the aims of the new health policy.

1.3 Governmental Policies to curb Malnutrition

Many endeavours are being undertaken to reduce the scourge of malnutrition especially in less developed countries where the burden is high. The Millennium Development Goal (MDG) 4, targets a reduction in the under-five mortality rate between the years 1990 and 2015 (WHO, 2013). Thus more holistic and integrated strategies have been developed which aims at maximum benefits to the most susceptible groups which are mainly children (under-five).

A major step in this stead has been the establishment and promotion of exclusive breast feeding and adequate nutritional diets for children less than five years old. The World Health Organisation (WHO) and the United Nations Children's Fund (UNICEF) in 1979 advised, at a breast-feeding programme, Exclusive Breast Feeding (EBF) for a period of 4-6 months. An expert committee in 2001, however, upon assessment of the efficiency of this programme then recommended that an EBF period for 6 months must be strictly adhered to, for optimal nutritional status of the child. Thus implying that no other liquids or complementary foods with the exception of undiluted drops or syrups consisting of vitamin and mineral supplements or medicines are given to the child at this stage. Also, during this period water is not permitted to be given to the child. (Gupta et al, 2002) and (Dewey et al, 1999).

The ministry of Health, Ghana, in collaboration with UNICEF, have recommended EBF for the first 6 months of the infant's life to mothers and caregivers. Though over 95% of children under five in Africa are currently practising EBF, effectiveness has not been achieved. This is because many mothers and caregivers feed their babies with water and other liquids alongside breast milk, thus contributing to a very low rate of EBF in Ghana and West Africa (Chinebuah-Aidam et al, 2004). In 2005 a strategy, High Impact and Rapid Delivery (HIRD) was introduced in Ghana, to help prevent avoidable deaths which results from ill health resulting from infections and malnutrition among children under-five years. This policy was implemented by the Ghana Health Services (GHS) and the Teaching Hospitals acting within the policy framework of the Ministry of Health (MOH), Ghana. (MOH, Ghana). The High-Impact Rapid-Delivery (HIRD) approach is a strategy to reduce maternal and child mortality. The HIRD approach combines the key principles of vision and data-driven methods to develop a plan for rapid scale up to attain universal (at least 90%) coverage of key priority cost effective interventions, which have been proven to have a high impact on maternal and child mortality. The HIRD initiative became necessary following the success of the Accelerated Child Survival and Development (ACSD) strategy in the Upper East region and the realization that unless there is accelerated pace in the reduction of maternal and ~~mortality~~ rate, Ghana will not achieve MDG 4 and 5.

Other key sectors directly involved in addressing nutrition issues include the:

1. Women in Agriculture Development Directorate (WIADD) of the Ministry of Food and Agriculture (MOFA) responsible for promoting food utilisation aspects

of MOFA through education, sensitization, awareness creation and advocacy.

2. Ghana Education Service (GES) of the Ministry of Education responsible for school feeding and health education programmes among others.

3. Children's Department of the Ministry of Women and Children's Affairs (MOWAC) responsible for child right promotion, child protection, and early childhood care and development (child health including nutrition through: education, sensitization, awareness creation and advocacy; child welfare; registration; etc.).

4. Department for Community Development of the Ministry of Local Government, Rural Development and Environment (MLGRDE) charged with the promotion of rural development issues, including the welfare and development of vulnerable groups in the rural and urban areas with focus on women and children; and

5. Department of Social Welfare of the Ministry of Manpower Development, Youth and Employment charged with promotion of the welfare of children, the youth and other vulnerable groups. (Ghartey, 2010)

1.3.1 Profile of Ghana

Ghana is a Sub-Saharan country on the West coast of Africa which is bordered by Togo on the east, Cte d'Ivoire on the west, Burkina Faso to the north and the Gulf of Guinea to the south. Ghana lies between longitudes and latitudes, Degrees Minutes Seconds: N 8 12' 36.2052", W 1 6' 36.0288", respectively. There are 10 regions in the country which are Northern, Upper East, Upper West, Brong-Ahafo, Ashanti, Eastern, Western, Central, Volta, and Greater Accra Regions. There are now 216 Administrative District Assemblies across the 10 regions in the country with 46 of them being added to the already existing 170 in 28 June

2012. (GhanaDistricts, 2013)

The population of Ghana is estimated to be 24,652,402 (July 2012). The population growth rate in Ghana is estimated to be 2.2% (2012) that is the country's comparison to the world. The age structure in the country is as follows; 0-14 years: 38.9% (male 4,815,972/ female 4,773,335),15-64 years: 57.1% (male 6,889,967/ female 7,179,274) ,65 years and over: 4% (male 463,962/ female 529,892) (2012 est.) with median age being; total: 21.7 years male: 21.4 years female: 21.9 years (2012).

The birth rate of Ghana in comparison to the world is at 32 births/ 1000 population, with the death rate being 7.7deaths/1000 population. The proportion of urban population in the country is 51% of total population (2010) and the rate of urbanization is 3.4% annual rate of change (2010-2015 est.). The sex ratio at birth in the country is 1.03 male(s)/female, under 15 years: 1.02 male(s)/female, 15-64 years: 1 male(s)/female, 65 years and over: 0.82 male(s)/female and total population: 1 male(s)/female. As of 2012, life expectancy at birth is about 61.45 years; 60.22 years for males and 62.73 for females with infant mortality rate being 40.9 deaths/1000 live births; males with 45.1/1000 live births and 36.7 deaths/1000 live births for females. The total fertility rate is about 4.115 per woman.(CIA, 2013)

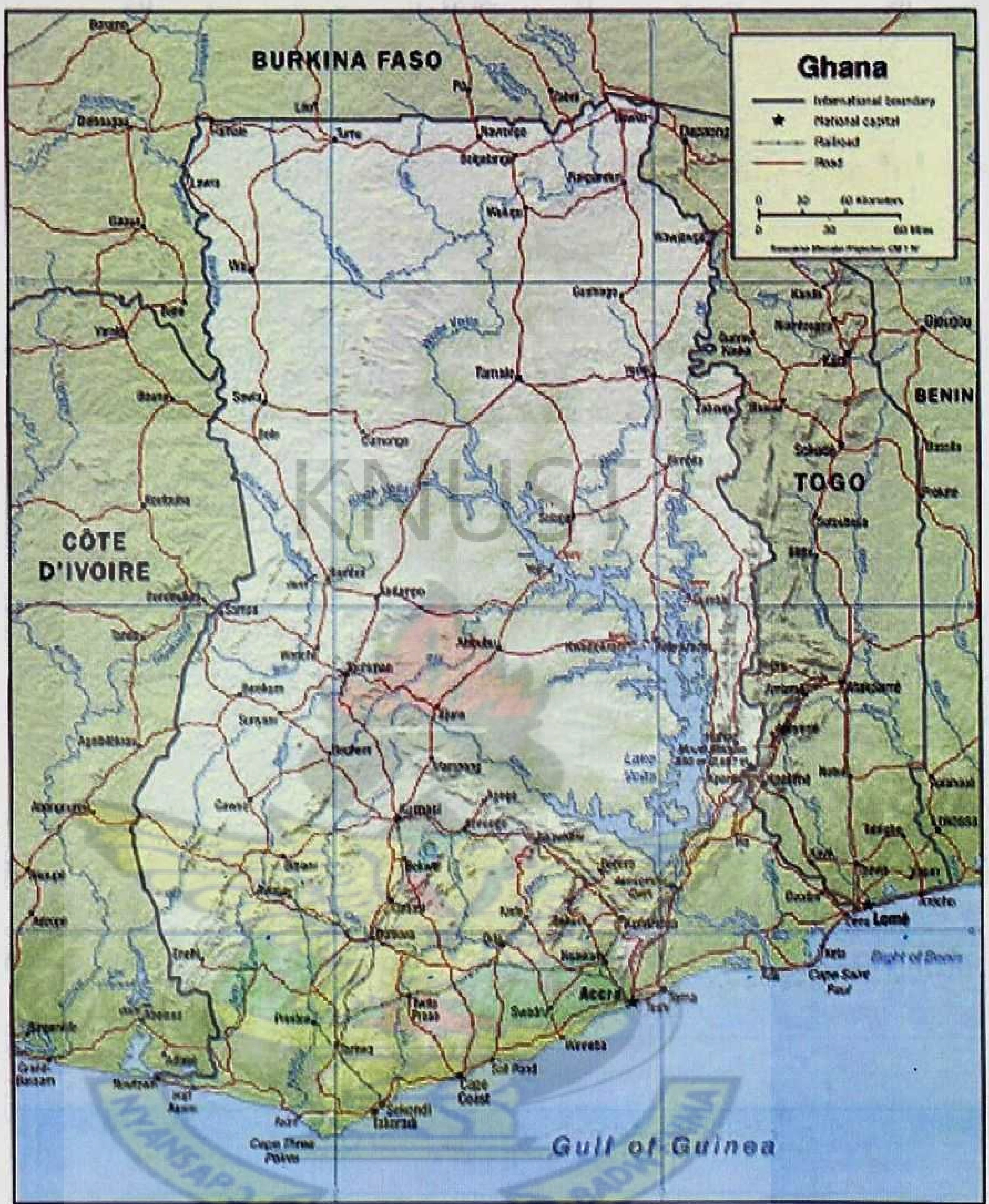

Figure 1.1: Source:www.lib.utexas.edu/maps/africa/ghana

1.4 Problem Statement

Malnutrition persists to be a nagging problem in Ghana. The most affected regions by malnutrition has been upper east, upper west and the northern regions. About 1.2 million people, representing 5 percent of Ghana's population, are food insecure. Thirty four percent (34%) of the population are in Upper West region, followed by Upper East with 15% and Northern region with 10%, amounting to approximately 453,000 people. This has been the case since poverty is high in these regions of the country. (MOFA, 2013).

Many factors can however be used to assess malnutrition, and many has been used over the years. Prominent among these is the conceptual frame work for malnutrition by UNICEF, which has been in use for the past 25 years. This conceptual framework was however modified in 2008 based on a research; Global, regional, and national causes of child mortality in 2008: A systematic analysis. (Black et al, 2010).

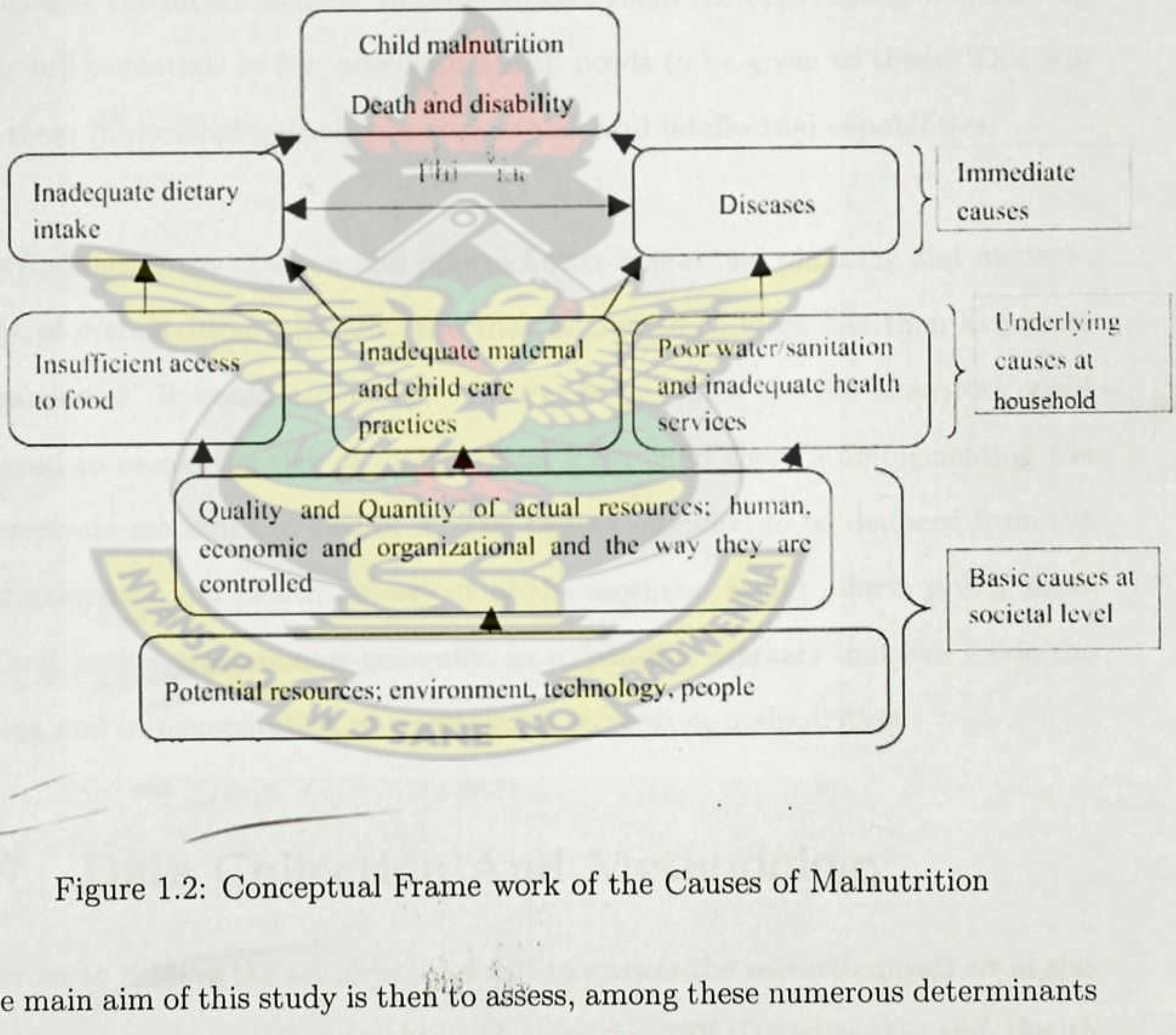

Figure 1.2: Conceptual Frame work of the Causes of Malnutrition

The main aim of this study is then to assess, among these numerous determinants of malnutrition, the latent or underlying construct which determines malnutrition in Ghana.

1.5 Objectives of the Research

The objectives of this study are as follows;

1. To identify the latent factors among the indicator variables and explain the interdependence that exists between them.

2. To establish the relationship among these latent factors or covariates identified.

1.6 Justification Of The Researech

The greatest asset of any country is the children that are born, as they will grow to become the future leaders. In order to give them the opportunity of achieving their full potentials in life, adequate health needs to be given to them. This will aid them in developing their physical, social, and intellectual capabilities.

The findings of the study would inform better contextual planning and management of malnutrition generally, and that related to children less than five years in particular. It would provide the framework by which specific indicators could be used to assess the risk of malnutrition for a child thereby implementing the appropriate measures to curtail it. The factors intended to be deduced from the characteristics and health behaviour of the mothers, would inform policy makers and health professionals generally, as to possible markers that can guide the design and implementation of intervention to prevent malnutrition.

1.7 Data Collection And Methodology

In order to achieve the set objectives and to answer the research questions of this study, a secondary data was obtained from the 2008 Demographic and Health Survey (DHS) conducted in Ghana. DHS has earned a worldwide reputation for collecting and disseminating accurate, nationally representative data on fertility, family planning, maternal and child health, gender, HIV/AIDS, malaria, and nutrition. Anthropometric data on age, height and weight (among other variables) collected in the 2008 GDHS permit the measurement and evaluation of the nutritional status of young children in Ghana. Variables relating to malnutrition in children to be considered in this study will include Child-level characteristics such as age, sex, duration of breastfeeding, size at birth; maternal characteristics such as education, mother's age at birth, birth interval, marital status, use of health services, occupation and finally household-level characteristics such as wealth, type of toilet facility, access to safe water, number of under-five children in the household, region and urbanization.

1.8 Scope And Limitation.

The method of Factor Analysis will be used in order to determine the underlying factors which engender these variables which relates to malnutrition in children. These factors will then be fitted in model which will seek to determine the presence of malnutrition in children under-five. Data for this study will be based on the 2008 Demographic and Health Survey (GDHS) conducted in Ghana. The children's recode which represents the entire data which was collected by the Demographic and Health Survey (GDHS) will be used in the analysis in order to determine the various latent factors that explain malnutrition. For the 2008 GDHS, the nutritional status of children is calculated using new growth standards published by the World Health Organisation (WHO) in 2006. These new growth standards were generated using data collected in the WHO Multicenter Growth Reference Study (WHO, 2006). Each of the three nutritional status indicators, height-for-age, weight-for-height, and weight-for-age, is expressed in standard deviation units from the median of the WHO Child Growth Standards. Statistical Product and Service Solutions (SPSS) will be used in the analysis.

1.9 Thesis Organization

This thesis is organized into five chapters. The first chapter provides the introduction which consists of background of the study, research problem, research question, research objective, justification, scope and limitation of the study, data analysis and data collection. The second chapter consists of a literature review. The methodology consists of data sources, sampling procedures, unit analysis, selected variables and method of analysis will be done in the third. The fourth covers the main analysis and data collection. Fifth chapter provides the findings, conclusion and recommendations.

Chapter 2

Literature Review

2.0 Introduction

This chapter seeks to describe Factor analysis, outline some of the principal uses of this technique and review research works conducted using this statistical technique by different authors. There will also be reviews of the concept of malnutrition, its effect on the society, economy and other health related factors affecting children under five years.

2.1 Factor Analysis

In order to explain the complex variety associated with inter-connections of social and international relations, there has been the proposal of thousands of variables. With these variables comes an, almost, equal number of hypotheses and theories linking these variables that has been suggested. The few basic variables and propositions central to understanding remain to be determined. The systematic dependencies and correlations among these variables have been charted only roughly, if at all, and many, if not most, can be measured only on presence-absence or rank order scales. To be able to take the data on any one variable at face value is to beg questions of validity, reliability, and comparability. With these entangled behaviour, unknown interdependencies, masses of qualitative and quantitative variables, and bad data, many social scientists are turning toward factor analysis to uncover major social and international patterns. (Adelman et al. 1965).

Factor analysis can simultaneously manage over a hundred variables, compensate for random error and invalidity, and disentangle complex interrelationships into their major and distinct regularities. (Rummel.1970). Factor analysis originated in psychometrics, and is used in behavioral sciences, social sciences, marketing, product management, operations research, and other applied sciences that deal with large quantities of data. Factor analysis was invented nearly 100 years ago by psychologist Charles Spearman, who hypothesized that the enormous variety of tests of mental ability–measures of mathematical skill, vocabulary, other verbal skills, artistic skills, logical reasoning ability, etc.,could all be explained by one underlying "factor" of general intelligence that he called 'g'. He hypothesized that if g could be measured and you could select a subpopulation of people with the same score on g, in that subpopulation you would find no correlations among any tests of mental ability. In other words, he hypothesized that g was the only factor common to all those measures. (Thomson 1947)

What factor analysis does is this: it takes thousands and potentially millions of measurements and qualitative observations and resolves them into distinct patterns of occurrence. It makes explicit and more precise the building of fact-linkages going on continuously in the human mind.

There are many statistical methods which can be used to study the relation between variables. However, Factor analysis is different; it is used to study the patterns of relationship among many variables, with the goal of discovering something about the nature of the latent variables that affect them, even though those latent variables were not measured directly. Therefore solutions obtained from the technique of factor analysis are necessarily more hypothetical, implicit and tentative than is true when variables are observed directly. The inferred latent variables are called factors. (Darlington, et al. 1973).

In Factor Analysis, the shared variance of a variable during factor extraction is separated from its unique variance and error variance to reveal the fundamental/latent factor structure; therefore only common variance appears in the solution. However, the use of Principal components analysis does not discriminate between shared and unique variance. Therefore in the case when factors are uncorrelated and communalities are moderate, Principal components analysis can produce inflated values of variance accounted for by the components (Gorsuch, 1997; McArdle, 1990).

Factor analysis can be carried out in order to explore a content area, structure a domain, map unknown concepts, classify or reduce data, reveal causal relationships, screen or transform data, define relationships, test hypotheses, formulate theories, control variables, or make inferences. (Rummel, 1967) A typical factor analysis suggests answers to major questions such as;

1. How many different factors are needed to explain the pattern of relationships among these variables?
2. What is the nature of those factors?
3. How well do the hypothesized factors explain the observed data?

Next some related research works undertaken by other authors will be reviewed.

Wolff and Wolff (1995), in an article investigated the relationship between maternal diet and infant birth weight. Principal components means of factor extraction was used in factor analysis to determine the structure of the maternal eating patterns. A total of 549 Mexican American mothers were sampled and a dietary criterion from the United States Hispanic Health and Nutrition Examination Survey was used to obtain data for the study. Results from Factor Analysis revealed 7 distinct underlying eating patterns: nutrient dense, traditional, transitional, nutrient dilute, protein rich, high fat dairy and mixed dishes. Afterwards, a stepwise multiple regression analysis was used to classify these eating patterns associated with birth weight. Other variables aside the eating patterns were included in the regression analysis. These variables included body mass index, haemoglobin, gestational age at delivery, maternal age, infant gender, acculturation, marital status, income, education, and smoking during pregnancy. Results from Regression analysis showed that the nutrient dense (fruits, vegetables, low fat dairy, etc.) and protein rich (low fat meats, processed meats, and dairy desserts, etc.) eating patterns were associated with increased birth weight and that the transitional eating pattern (fats and oils, breads and cereals, high fat meats, sugar, etc.) was associated with decreased birth weight. Study findings suggest that the eating pattern methodology may be an appropriate tool for analyzing food frequency data in the investigation of diet and health relationships and for targeting dietary interventions.

Rikimaru et al. (1998), in the Princess Marie Louise Hospital, Accra, Ghana, a case-control study was conducted to examine the association between immunological variables and the diverse types and severity of malnutrition in Ghanaian children. 170 children, aged 8-36 month were sampled at the clinical ward and public health service section of the hospital: 61 healthy children, 49 moderately malnourished (underweight) children and 60 severely malnourished children that is, 19 kwashiorkor, 30 marasmus, and 11 marasmic kwashiorkor children. In the clinical observations, anthropometric measurements and blood sampling for biochemical analysis were taken in order to evaluate their nutritional and immunological status. Serum immunoglobulins (IgA subclasses, IgG subclasses and IgM), complements (C3 and C4) and lymphocyte subpopulations (T cells, B cells, CD4+, CD8+, NK cells and HLADR) were determined for the assessment of humoral and cell-mediated immunity. It was found out that Serum levels of IgA1, IgA2 and C4 tended to be higher in children who were severely malnourished than in healthy children, while serum level of C3 and the proportion of B cells were significantly lower in the severely malnourished children than in the normal children ($P < 0.05$). The results of Factor analysis revealed that C3 levels were positively correlated with a factor which was strongly associated with weight-for-height z-score and biochemical indicators for evaluating protein nutrition. In addition, IgA2, IgG1 and IgM levels were positively correlated with a factor which was associated with C-reactive protein.

Cheah et al. (2009) in a study to develop a questionnaire and validate the factors associated with malnutrition among children in rural Kelantan, Malaysia applied the method of factor analysis. They identified the problem of factors associated with or contributing to malnutrition among children to be are diverse, multisectoral and include interrelated biological, social, cultural and economic factors. In order to ascertain these factors in a given population an accurate and reliable questionnaire was designed initially. Their study was initially based on an already existing conceptual framework of malnutrition in children and was conducted in two phases. A semi-structured questionnaire was administered initially for 20 health workers who provide direct care for the malnourished study children in rural Kelantan, Malaysia. The results from this initial findings aided in a theoretical framework being generated to assist the development of quantitative questionnaire of 17 variables. The questionnaire was then administered to 295 children and their parents/caregivers, Exploratory factor analysis then revealed two factors: environmental and behavioral with composite reliability of 0.70, and 0.74, respectively.

Newby et al. (2006), in a study to find the long-term stability of food patterns in Swedish women, used factor analysis among 33,840 women participating in the Mammography Cohort. Afterwards, comparison of the factor solutions from confirmatory factor analysis with those derived by use of exploratory factor analysis. Dietary patterns was assessed by a food frequency questionnaire which was developed from 1987 and 1997, and food patterns were derived by the use of exploratory and confirmatory factor analysis. The results identified four dietary patterns: Healthy, Western/Swedish, Alcohol, and Sweets. Correlations between confirmed food patterns in 1987 and 1997 were 0.37 for the Healthy pattern, 0.27 for the Western/Swedish pattern, 0.54 for the Alcohol pattern, and 0.44 for the Sweets pattern . Patterns derived by the use of exploratory factor analysis were strongly associated with those derived by the use of confirmatory factor analysis . The patterns derived in this study were similar to those derived in other studies, indicating reproducibility of food patterns across populations.

Flood et al. (2008), investigated associations between dietary patterns and colorectal cancer, using factor analysis in middle-aged Americans. Diet has long been suspected as an aetiological factor for colorectal cancer, but studies of single foods and nutrients have provided inconsistent results. Diet was assessed among 293615 men and 198767 women in the National Institutes of Health-AARP Diet and Health Study. Principal components factor analysis identified 3 primary dietary patterns: a fruit and vegetables, a diet foods, and a red meat and potatoes pattern. State cancer registries identified 2151 incident cases of colorectal cancer in men and 959 in women between 1995 and 2000.

Lioret et al. (2008), in a research, studied if lifestyle patterns, combining overall diet and physical activity, were associated with childhood Over Weight incidence, and if they were involved in the reverse association between socioeconomic status (SES) and Over Weight. A total of 748 French children between the ages 3?11 years from the 1998?1999 cross-sectional French Enqute Individuelle et Nationale sur les Consommations Alimentaires national food consumption survey, were sampled. Variables collected in the survey were, Weight and height, leisure time physical activity, SED (television viewing), and SES . Data was obtained from parents or children, who answered the questionnaires administered to them. Scores for lifestyle patterns were assessed with factor analysis and their relationship with Over Weight was explored by logistic regression analysis. Factor analysis identified 2 similar lifestyle patterns: 'snacking and sedentary' and 'varied food and physically active'. The snacking and sedentary pattern was positively correlated with Over Weight in very young children and partly mediated the negative association of SES to Over Weight. However, the second factor, 'varied food and physically active' pattern was inversely correlated with Over Weight in the eldest children only. A third pattern called 'big eaters at main meals' was derived in children aged between 7−11 years and was positively correlated with Over Weight.

2.2 Measurement of Nutritional Status among Children under five

The 2008 (Ghana Demographic Health Survey) GDHS organized in Ghana collected data on the nutritional status of children in the country by measuring the height and weight of all children under six years of age. Although data were collected for all children under age six, for purposes of comparability, the analysis is limited to children under age five. The aim of the collected measurements was the calculation of the ~~three~~ indices which tells of the nutritional status of children; weight-for-age, height-for-age, and weight-for-height, all of which take age and sex into consideration.

The Weight measurements were taken from children by using lightweight, electronic Seca scales with a digital screen, which were designed and manufactured under the guidance of the United Nations Children's Fund (UNICEF). Height measurements were also taken by using a measuring board produced by Shorr Productions. However, Children with ages less than 24 months were measured lying down (recumbent length) on the board while standing height was measured for children older than 24 months.

In the 2008 GDHS, the nutritional status of children was calculated using new growth standards published by the World Health Organisation (WHO) in 2006. These new growth standards were come about by generating data collected in the WHO Multicentre Growth Reference Study (WHO, 2006). Accordingly, each of these three nutritional status indicators mentioned above were expressed in standard deviation units from the median of the WHO Child Growth Standards. The indices are not comparable with those based on the previously used NCHS/CDC/WHO Reference.

The three indices, height-for-age, weight-for-height, and weight-for-age, respectively, provides divergent information about growth and body composition that is used to evaluate nutritional status. The height-for-age index is an indicator of linear growth retardation and cumulative growth deficits. Children are considered short for their age (stunted) and are chronically malnourished if it is found put that their height-for-age Z-score is below minus two standard deviations (-2 SD). However, children whose height-for-age Z-score is below minus three standard deviations (-3 SD) are considered severely stunted. Stunting in children is reflective of the inability of the body to receive adequate nutrition over a long period of time and is also affected by repeated and chronic illness. Therefore, Height-for-age would represent the long-term effects of malnutrition in a population and is not sensitive to recent, short-term changes in dietary intake.

The weight-for-height index is an indicator of the body's mass in relation to body height or length and also describes current nutritional status. Children are considered thin (wasted) and are acutely malnourished if it is found out that their weight-for-height Z-score is below -2 SD. Wasting in children is reflective of the body's inability to receive adequate nutrition in the period immediately preceding the (2008 GDHS) survey and may be the result of inadequate food intake or a recent occurrence of illness causing loss of weight and the beginning of malnutrition. However, children whose weight-for-height is below -3 SD are considered severely wasted.

Weight-for-age is conversely a combined index of height-for-age and weight-for-height. This index takes into account both acute and chronic malnutrition. Children are classified as underweight if their measured weight-for-age index is below -2 SD. However, children whose weight-for-age is below -3 SD are considered severely underweight. (GSS et al. 2008)

2.3 Levels of Malnutrition in Ghana

According to the results from the 2008 GDHS, 28 percent of children under five are stunted (below -2 SD), with 10 percent being severely stunted (-3 SD). Children 18-23 months old (40 percent) are likely to be more stunted than children who are less than 6 months old (4 percent). Male children (30 percent) are also slightly more likely to be stunted than female children (26 percent). There is a reduction of the extent of stunting as the birth interval and size at birth increase, and as the mother's Body Mass Index (BMI) increases. The level of stunting is higher in the rural areas (32 percent) than in the urban areas (21 percent). Stunting also varies by region; with percentage of stunting being highest in the Eastern and Upper East regions (38 and 36 percent, respectively) and lowest percentage in the Greater Accra region (14 percent). Stunting also decreases as mother's level of education and wealth quintile increase.

Overall, 9 percent of children under five are wasted, with 2 percent severely wasted according to the 2008 GDHS. Wasting is most prevalent among children within age 6-8 months (29 percent) and is lowest among children age 48-59 months (3 percent). The level of wasting does not vary much with sex, birth interval, or urban-rural residence. There is a reduction of the extent of wasting as the size at birth increases and mother's nutritional status improves. Wasting is also more prevalent in the Upper West (14 percent), Northern (13 percent) and Central (12 percent) regions than all the remaining other regions. There is also a reduction in the incidence of wasting as the mother's level of education and wealth quintile increase.

For children whose weight-for-age is below minus two standard deviations (-2 SD) from the median of the reference population are considered to be underweight. This measure also reflects the effects of both acute and chronic malnutrition. From the results of the 2008 GDHS, 14 percent of children are underweight, with 3 percent classified as severely underweight. The age group distribution of underweight children is as follows; children age 18-23 months (19 percent), and followed by those age 9-11 months (18 percent). Male children (15 percent) are also slightly more likely to be underweight than female children (12 percent). There is a reduction of underweight children as birth interval, size at birth and mother's nutritional status increase. Furthermore, the percentage children living in rural areas (16 percent) are more likely to be underweight as compared to those in the urban areas (11 percent). The country regional proportion of underweight children ranges from 7 percent in Greater Accra to 27 percent in the Upper East region. Children born to mothers with little or no education are substantially more likely to be underweight than children of more educated women. For example, the proportion of underweight children born to women with no education was 17 percent, compared with 7 percent among children born to women with secondary education or higher. Similarly, children from households in the two wealthiest quintiles are the least likely to be underweight (8-9 percent).

2.4 Socio-Economic Factors on nutritional status of children

Earlier studies on nutritional status among children have discovered that there are significant links between maternal nutrition knowledge and child nutritional status (Ruel et al, 1992). Socio-cultural factors and patterns of parents, especially mothers, have been identified by some scientist (Ojofeitimi et al, 2003; Agnarsson, et al, 2001) as related to the nutritional status of children. In a recent research on the socio-cultural influences on feeding practices of women in Kwa-Zulu Natal, showed that mother's age has the most influence on the food choices for children less than five years. Older mothers feeding practices to their children tend to be more independent than those of a younger age (Thairu et al, 2005). Mckeever and Miller (2004), also found out that a child's nutritional status is enhanced by the mother's background characteristics, including age, employment status and educational status.

Appoh and Krekling, 2005, remarked that there was substantive evidence linking maternal nutritional knowledge and socio-economic status influences the nutritional status of their children in the Volta region of Ghana. Among 110 mothers interviewed for the study, there was also a strong association between the marital status of mothers and the nutritional status of their children upon their analysis. Furthermore, maternal knowledge with its incumbent practices on breastfeeding of children was a significant indicator for the nutritional status of the child. Therefore, factors such as marital status, educational status, and socio-economic status had a significant association with the nutritional status of the child.

Chapter 3

Methodology

Introduction

This chapter discusses the fundamental theory of Factor Analysis with regards to its definition, model and the method of analysis of the current data to arrive at the objective.

3.1 Basic Definition

Factor Analysis as statistical technique fundamentally looks to identify, the relationship that exists between covariance among several variables in terms of a few underlying or latent but unobservable, random quantities called factors. Suppose variables can be grouped by their correlations, then a particular group would consist of variables that correlate highly among themselves, but have relatively small correlations with variables in a different group. Then it can be envisioned that each group of variables represents a single underlying or latent factor, that is responsible for the group of correlated variables. For example, a class of students in basic school can be put into 2 groups. The first group is based on their test scores in classics, French, English, mathematics, and music and an underlying "intelligence" factor can be proposed. A second group of variables ,physical-fitness, might correspond to their performance on track and field events. This is the type of structure that factor analysis seeks to confirm. (Johnson and Wichern, 2002) Factor analysis can be thought of as an extension of principal component analysis. Both of these analyses can be viewed as statistical techniques to approximate the covariance matrix Σ. However the approximation based on the factor analysis model is more elaborate. The primary question in factor analysis is whether the data are consistent with a prescribed structure. The basic differences between Principal component analysis and Factor Analysis are:

1. Principal components are defined as linear combinations of the original variables, whereas in factor analysis, the original variables are expressed as linear combinations of factors that caused the original variables.

2. In Principal component analysis, we explain a large part of the total variance of the variables, while in Factor Analysis we seek to account for the covariances or correlations among the variables (Rencher and Christensen, 2012).

In factor analysis the variables $y_1, y_2, ..., y_p$ are represented as linear combinations of a few other groups $f_1, f_2, ..., f_m$,$(m < p)$ called factors. .These factors are the underlying constructs or latent variables that engender the variables . Like the original variables, the factors vary from individual to individual but unlike the variables, the factors cannot be measured or observed. If the original variables $y_1, y_2, ..., y_p$ are at least moderately correlated, the basic dimensionality of the system is less than p. The goal of factor analysis is to reduce the redundancy among the variables by using a smaller number of factors. Usually factor analysis uses the correlation matrix to determine the latent factors. In some cases where the correlation matrix does not have such a simple pattern, factor analysis will partition the variables into clusters.(Rencher and Christensen, 2012)

3.2 Orthogonal Factor Model

Factor Analysis is basically a one-sample procedure. Assumption is made of a random sample $y_1, y_2, ..., y_p$ from a homogeneous population with mean μ and covariance matrix Σ

The Factor Analysis model expresses each variable as a linear combination of underlying common factors $f_1, f_2, ..., f_m$ with an accompanying error term to account for that part of the variable that is unique, that is not common with the other variables.

For $y_1, y_2, ..., y_p$ in any observational vector , the model is as follows:

$$\begin{aligned}
y_1 - \mu_1 &= \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{1m} f_m + \epsilon_1 \\
y_2 - \mu_2 &= \lambda_{21} f_1 + \lambda_{22} f_2 + \cdots + \lambda_{2m} f_m + \epsilon_2 \\
&\vdots \\
y_p - \mu_p &= \lambda_{p1} f_1 + \lambda_{p2} f_2 + \cdots + \lambda_{pm} f_m + \epsilon_p
\end{aligned} \tag{3.1}$$

Where $(m < p)$ ideally, otherwise we have not achieved a parsimonious description of the variables as functions of a few underlying factors: We note that the $f's$ are the random variables that engender the variables $y's$. λ_{ij} are called loadings and serve as weights showing how each y_i individually depends on the $f's$. With appropriate assumptions, λ_{ij} indicates the importance of the j^th factor to the i^th variable y_i. The larger loadings relate the $f's$ to the corresponding y_i. After estimating the $(\lambda_{ij})'s$ it is hoped they will partition the variables into groups corresponding to factors.

The observable random vector Y with p components, has mean U and covariance matrix Σ. The factor model postulates that Y is linearly dependent upon a few unobservable random–variables vector F , which is called a vector of common factors. Therefore the p additional sources of variation also form an error vector

E. The factor analysis model is then;

$$\underset{(p\times 1)}{Y-U} = \underset{(p\times m)}{\Lambda}\underset{(m\times 1)}{F} + \underset{(p\times 1)}{E} \tag{3.2}$$

Where, $Y = (y_1, y_2, \dots, y_p)\prime$,

$U = (\mu_1, \mu_2, \dots, \mu_p)\prime$,

$F = (f_1, f_2, \dots, f_m)\prime$,

$E = (\epsilon_1, \epsilon_2, \dots, \epsilon_p)\prime$

and $\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2m} \\ \vdots & \vdots & \dots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pm} \end{pmatrix}$

3.2.1 Assumptions of the model

For $j = 1, 2, \dots, m$

$E(f_j) = 0$, $var(f_j) = 1$ and $cov(f_j, f_k) = 0$, for all $j \neq k$

The assumptions for ϵ_i, $i = 1, 2, \dots, p$ are similar except that each ϵ_i must be allowed to have a different variance, since it shows the residual part of y_i that is not common with other variables. Therefore;

$E(\epsilon_i) = 0$, $var(\epsilon_i) = \varphi_i$, that is a specific variance, and $cov(\epsilon_i, \epsilon_k) = 0$, for all $i \neq k$

An assumption of $cov(\epsilon_i, f_j) = 0$ for all $i\&j$, is now made.

Now, $E(y_i - \mu_i) = 0$, thus we need $E(f_j) = 0$, $j = 1, 2, \dots, m$. The assumption os $cov(f_j, f_k) = 0$ is made for pasimony in expressing the $y's$ as functions of as few factors as possible.

Next, we assume that $var(f_j) = 1$, $var(\epsilon_i) = \varphi_i$, and $cov(\epsilon_i, f_j) = 0$. These results enables us to obtain a simple expression for the variance of y_i, which is;

$$var(y_i) = \lambda_{i1}^2 + \lambda_{i2}^2 + \cdots + \lambda_{im}^2 + \varphi_i \tag{3.3}$$

Note that the assumption $cov(\epsilon_i, \epsilon_k) = 0$ implies that the factors account for all the correlations among the $y's$, that is all that the variables have in common. Thus the emphasis in Factor Analysis is on the modelling of the covariances or correlations among the variables, $y's$.

3.2.2 Assumptions in Matrix and Vector Notations

The afore mentioned assumptions in the Factor Analysis model is made with a univariate data in mind. An extension of the already mentioned assumption is used in the Matrix and Vector notations.

1. $E(f_j) = 0$ for $j = 1, 2, \ldots, m$ becomes $E(F) = O$ (a zero matrix)
2. $var(f_j) = 1$ for $j = 1, 2, \ldots, m$ and $cov(f_j, f_k) = 0$, for all $j \neq k$, become $cov(F) = I$
3. $E(\epsilon_i) = 0$, for $i = 1, 2, \ldots, p$, becomes $E(E) = O$ (a zero matrix)
4. $var(\epsilon_i) = \varphi_i, i = 1, 2, \ldots, p$ and $cov(\epsilon_i, \epsilon_k) = 0$, for all $i \neq k$ become $cov(E) = \Psi$, that is a diagonal matrix with $\varphi_i, i = 1, 2, \ldots, p$ in the main diagonal. Therefore,
$$cov(E) = \Psi = \begin{pmatrix} \varphi_1 & 0 & \ldots & 0 \\ 0 & \varphi_2 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & \varphi_p \end{pmatrix}$$
5. $cov(\epsilon_i, f_j) = 0$, for all i & j becomes $cov(E, F) = O$, indicating a rectangular zero matrix containing the covariances of the factors $f's$ with the error terms $\epsilon' s$.

3.3 The simplified Structure

The covariance and variance of the variables $y_1, y_2, ..., y_p$ can be expressed in terms of a simplified structure. This will involve the $p \times m$ loadings λ_{ij} and the φ_i specific variances and therefore the equation Σ, will be expressed in terms of Λ and Ψ . Now, since the variances and the covariances of Y are not affected by μ , we have;

$$\Sigma = cov(Y) = cov(\Lambda F + E) \tag{3.4}$$

Because ΛF and E have zero correlation, the covariance matrix of their sum is the sum of their covariance matrices.

$$\begin{aligned} \Sigma &= cov(\Lambda F) + cov(E) \\ &= \Lambda cov(F)\Lambda' + \Psi \\ &= \Lambda I \Lambda' + \Psi \\ \Sigma &= \Lambda\Lambda' + \Psi \end{aligned} \tag{3.5}$$

Additionally the covariances of the variables $y_1, y_2, ..., y_p$ can be expressed with the $f's$ in terms of the $\lambda' s$. For instance consider $cov(y_1, f_2)$. From (3.1) we deduce the equation $y_1 - \mu_1 = \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{1m} f_m + \epsilon_1$. f_2 is uncorrelated with all the other $f_j's$ since $cov(f_j, f_k) = 0$, $j \neq k$. f_2 is also not correlated with ϵ_1 since $cov(\epsilon_i, f_j) = 0$ for all $i \& j$. Then we have;

$$cov(y_1, f_2) = E[(y_i - \mu_1)(f_2 - \mu_{f_2})]$$

since $E[f_j] = \mu_{f_j} = 0$

$$
\begin{aligned}
cov(y_1, f_2) &= E[(\lambda_{11}f_1 + \lambda_{12}f_2 + \cdots + \lambda_{1m}f_m + \epsilon_1)f_2] \\
&= E[(\lambda_{11}f_1f_2 + \lambda_{12}{f_2}^2 + \cdots + \lambda_{1m}f_mf_2)] \\
&= \lambda_{11}cov(f_1, f_2) + \lambda_{12}var(f_2) + \cdots + \lambda_{1m}cov(f_m, f_2) \\
cov(y_1, f_2) &= \lambda_{12}
\end{aligned}
$$

since,

$$E[(f_1 - E(f_1))(f_2 - E(f_2))] = E(f_1, f_2) = cov(f_1, f_2)$$

In general, the loadings themselves represent covariances of the variables with the factors. That is,

$$cov(y_i, f_j) = \lambda_{ij} \tag{3.6}$$

$\forall\ i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, m$

With λ_{ij} being the ij^{th} element of Λ (from equation 3.2). The vector Notation of (3.6) is

$$cov(Y, F) = \Lambda \tag{3.7}$$

In (3.3) we have a partitioning of the variance of y_i into a component due to the common factors called Communality and a component unique called the specific variance.

$$
\begin{aligned}
var(y_i) = \sigma_{ii} = \lambda_{i1}^2 + \lambda_{i2}^2 + \cdots + \lambda_{im}^2 + \varphi_i \\
= h_i^2 + \varphi_i
\end{aligned} \tag{3.8}
$$

$$= communality + specific\ variance$$

3.4 Nonuniqueness of Factor Loadings

Factor loadings in the model $Y - U = \Lambda F + E$ can be multiplied by an orthogonal matrix without taking away its ability to produce or replicate the covariance matrix $\Sigma = \Lambda\Lambda' + \Psi$.

For this property, we let T be an arbitrary orthogonal matrix where $TT' = I$. This will be then be inserted in the equation (3.2).

$$Y - U = \Lambda TT'F + E \tag{3.9}$$

Next a relationship between T and Λ, and T' and F are established, then the equation above becomes;

$$Y - U = \Lambda^* F^* + E \tag{3.10}$$

where:

$$\Lambda^* = \Lambda T \tag{3.11}$$

$$F^* = T'F \tag{3.12}$$

Next Λ in (3.4) is replaced with $\Lambda^* = \Lambda T$, and obtain

$$\Sigma = \Lambda^* \Lambda^{*\prime} + \Psi \tag{3.13}$$

But;

$$\begin{aligned}
\Sigma &= \Lambda T(\Lambda T)' + \Psi \\
&= \Lambda TT'\Lambda' + \Psi \\
&= \Lambda I \Lambda' + \Psi \\
&= \Lambda\Lambda' + \Psi
\end{aligned}$$

Therefore, the new factor loadings $\Lambda^* = \Lambda T$ also reproduce a covariance matrix just as the previous factor loadings Λ does. An equivocation of the two covariance matrices can then be made in this form,

$$\Sigma = \Lambda^*\Lambda^{*\prime} + \Psi = \Lambda\Lambda' + \Psi \tag{3.14}$$

Therefore proving that factor loadings are not unique.

The new factors will also satisfy similar assumptions mentioned earlier. That is,

1. $E(F^*) = O$
2. $cov(F^*) = I$
3. $cov(F^*, E) = O$

The communalities as seen in $h_i^2 = \lambda_{i1}^2 + \lambda_{i2}^2 + \cdots + \lambda_{im}^2$,$i = 1, 2, \ldots, p$ are also not affected by the new factor loadings, $\Lambda^* = \Lambda T$. We notice the communalities h_i^2 is the sum of the squares of the ith row of Λ.
Let χ_i' represent the elements in the ith row of Λ, then the sum of squares can be expressed in vector notation as $h_i^2 = \chi_i'\chi_i$. The ith row of with the new factor loadings ($\Lambda^* = \Lambda T$) will now be $\chi_i^{*\prime} = \chi_i'T$ and the corresponding communality is

$$h_i^{*\prime} = \chi_i^{*\prime}\chi_i^* \tag{3.15}$$

where;

$$\begin{aligned} &= \chi_i'TT'\chi_i \\ &= \chi_i'\chi_i \\ &= h_i^2 \end{aligned}$$

Therefore the communalities also remain unchanged for the new loadings. It is worth noting that $h_i^2 = \lambda_{i1}^2 + \lambda_{i2}^2 + \cdots + \lambda_{im}^2$ is the distance between the origin to the point $\chi_i' = (\lambda_{i1}, \lambda_{i2}, \ldots, \lambda_{im})$ in the m-dimensional space of the factor loadings. Since the distance $\chi_i'\chi_i$ is the same as $\chi_i^{*\prime}\chi_i^*$, the points χ_i^* are rotated from the points χ_i.

3.5 Estimation of Loadings and Communalities

Methods for estimating factor loadings and the common variance (communalities) have been proposed by various proponents of the use of factor analysis. Four of such techniques will be discussed in the following subsections.

3.5.1 Principal Component Method

In the Principal Component method of estimating factor loadings, a sample covariance matrix S is extracted from a random sample $y_1, y_2, \ldots, y_p$ and then an estimator $\hat{\Lambda}$ is also obtained. This estimator seeks to approximate the fundamental expression, $\Sigma = \Lambda\Lambda' + \Psi$ with S in place of Σ ;

$$S \cong \hat{\Lambda}\hat{\Lambda}' + \Psi \tag{3.16}$$

Next, the covariance of the error term is omitted and is factored into $S \cong \hat{\Lambda}\hat{\Lambda}'$. This enables the use of the spectral decomposition of the sample covariance S in the form;

$$S = CDC' \tag{3.17}$$

That is, C is an orthogonal matrix constructed with normalized eigenvectors $(c_i')c_i = 1$ of the sample covariance matrix S .D is a diagonal matrix with the eigenvalues of S ; $\theta_1, \theta_2, \ldots, \theta_p$ in its main diagonal.

$$D = \begin{pmatrix} \theta_1 & 0 & \ldots & 0 \\ 0 & \theta_2 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & \theta_p \end{pmatrix} \tag{3.18}$$

To complete the factoring of CDC' in (3.17) into the required form $\hat{\Lambda}\hat{\Lambda}'$, the matrix D can be revised. This is because the eigenvalues are derived from a positive semidefinite sample covariance matrix S , therefore all the $\theta_i{}'s$ are positive or zero;

$$D = D^{1/2}D^{1/2}$$

where

$$D^{1/2} = \begin{pmatrix} \sqrt{\theta_1} & 0 & \dots & 0 \\ 0 & \sqrt{\theta_2} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sqrt{\theta_p} \end{pmatrix}$$

Hence equation (3.16) becomes

$$\begin{aligned} S = CDC' &= CD^{1/2}D^{1/2}C' \\ &= (CD^{1/2})(CD^{1/2})' \end{aligned} \tag{3.19}$$

(3.19) is then of the form $S = \hat{\Lambda}\hat{\Lambda}'$. However, $CD^{1/2}$ is a $p \times p$ matrix but we require the estimate $\hat{\Lambda}$ to be a $p \times m$ matrix with $p > m$, in order to achieve the parsimonious objective of finding the factors. Therefore $\hat{\Lambda}$ cannot be defined to be $CD^{1/2}$. Given $\theta_1 > \theta_2 > \cdots > \theta_m$, that is the m largest eigenvalues, we can define $D_1 = diag(\theta_1, \theta_2, \dots, \theta_m)$ and which comprises the corresponding eigenvectors. Then the estimate of Λ can be estimated using the first m columns of $CD^{1/2}$, that is

$$\hat{\Lambda} = C_1 D_1^{1/2} = (\sqrt{\theta_1}c_1, \sqrt{\theta_2}c_2, \dots, \sqrt{\theta_m}c_m) \tag{3.20}$$

Where $\hat{\Lambda}$ is $p \times m$, that is C_1 is $p \times m$ and D_1 is $m \times m$. This can be illustrated for a structure of the $\hat{\lambda}_{ij}$ in the form of (3.20) with $p = 4$ and $m = 2$

$$\hat{\Lambda} = \begin{pmatrix} \hat{\lambda}_{11} & \hat{\lambda}_{12} \\ \hat{\lambda}_{21} & \hat{\lambda}_{22} \\ \hat{\lambda}_{31} & \hat{\lambda}_{32} \\ \hat{\lambda}_{41} & \hat{\lambda}_{42} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \end{pmatrix} \begin{pmatrix} \sqrt{\theta_1} & 0 \\ 0 & \sqrt{\theta_2} \end{pmatrix}$$

$$\begin{pmatrix} \sqrt{\theta_1}c_{11} & \sqrt{\theta_2}c_{12} \\ \sqrt{\theta_1}c_{21} & \sqrt{\theta_2}c_{22} \\ \sqrt{\theta_1}c_{31} & \sqrt{\theta_2}c_{32} \\ \sqrt{\theta_1}c_{41} & \sqrt{\theta_2}c_{42} \end{pmatrix} \tag{3.21}$$

The columns of $\hat{\Lambda}$ are relative to the eigenvectors of S, therefore the loadings on the jth factor are relative to coefficients in the jth principal component. The factors are then related to the first m principal components, and it would seem that similar interpretation would be given as Principal component Analysis. However the interpretation would be different after rotation of the factor loadings. To complete this approximation of S, we note that the ith element of $\hat{\Lambda}\hat{\Lambda}'$ is the sum of squares of the ith row of $\hat{\Lambda}$, that is $\hat{\lambda}_i'\hat{\lambda}_i = \sum_{j=1}^{m} \hat{\lambda}_{ij}^2$.

Therefore,

$$\hat{\varphi}_i = s_{ii} - \sum_{j=1}^{m} \hat{\lambda}_{ij}^2 \tag{3.22}$$

and write

$$S \cong \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi} \tag{3.23}$$

with $\hat{\Psi} = diag(\hat{\varphi}_1, \hat{\varphi}_2, \ldots, \hat{\varphi}_p)$

The variances on the main diagonal of S (3.23) are obtained exactly, only the off-diagonal covariances are approximates. Therefore, the sum of squares of the rows of S are the communalities, while that of the columns are the eigenvalues.

$$\hat{h}_i^2 = \sum_{j=1}^{m} \hat{\lambda}_{ij}^2 \tag{3.24}$$

(3.24) form the sum of squares of the ith row of $\hat{\Lambda}$. The sum of squares of the jth column of $\hat{\Lambda}$ is the jth eigenvalue of S :

$$\begin{aligned} \textstyle\sum_{i=1}^{p} \hat{\lambda}_{ij}^2 &= \textstyle\sum_{i=1}^{p} (\sqrt{\theta_j} c_{ij})^2 \qquad [by\ (3.21)] \\ &= \theta_j \textstyle\sum_{i=1}^{p} c_{ij}^2 \\ &= \theta_j \end{aligned} \tag{3.25}$$

That is, $\sum_{i=1}^{p} c_{ij}^2 = 1$, normalized eigenvectors have length of 1.

From equations (3.22) and (3.24), variance of the ith variable can be further partitioned into two distinct parts. A part due to the factors (communalities) and another part due to the variable only:

$$\begin{aligned} s_{ii} &= \hat{h}_i^2 + \hat{\varphi}_i \\ &= \hat{\lambda}_{i1}^2 + \hat{\lambda}_{i2}^2 + \cdots + \hat{\lambda}_{im}^2 + \hat{\varphi}_i \end{aligned} \tag{3.26}$$

Therefore the jth factor adds $\hat{\lambda}_{ij}^2$ to the variance of the ith variable . For that of the population variance Ş the contribution of the jth factor is the trace of Ş

$$tr(\text{Ş}) = s_{11} + s_{22} + \cdots + s_{pp}$$

We then obtain variance to jth factor as

$$\sum_{i=1}^{p} \hat{\lambda}_{ij}^2 = \hat{\lambda}_{1j}^2 + \hat{\lambda}_{2j}^2 + \cdots + \hat{\lambda}_{pj}^2 \tag{3.27}$$

(3.27) then forms the sum of squares of loadings in the jth column of $\hat{\lambda}$ which is the same as θ_j, the jth eigenvalue. The proportion of total sample variance to the jth factor is;

$$\frac{\sum_{i=1}^{p} \hat{\lambda}_{ij}^2}{tr(\text{Ş})} = \frac{\theta_j}{tr(\text{Ş})} \tag{3.28}$$

The main goal of parsimony in the variables which factor analysis seeks to achieve is most of the time based on reproducing the covariances or correlations other than the variances. The eigenvalues and eigenvectors of the correlation matrix

R of the variabes $y_1, y_2, \ldots, y_p$, are then used in place of the variance of the variables,Ş . This is usually the case when the variables are not commensurate and standard variables are used instead. The proportion of total sample variance to the jth factor will then be;

$$\frac{\sum_{i=1}^{p} \hat{\lambda}_{ij}^2}{tr(R)} = \frac{\theta_j}{P} \tag{3.29}$$

Where P represents the number of variables.

(3.22), $S \cong \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$ can be rewritten in a form to assess the fit of the factor analysis model.

$$E = S - (\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}) \tag{3.30}$$

Where E is the error matrix, which by definition, has zeros on the main diagonal but non-zero elements in the off-diagonals.

The following inequality gives a bound on the size of elements in the error matrix E:

$$\sum_{ij} \epsilon_{ij}^2 \leq \theta_{m+1}^2 + \theta_{m+2}^2 + \cdots + \theta_p^2 \tag{3.31}$$

That is, the sum of squared entries in the matrix $E = S - (\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi})$ is at most equal to the sum of squares of the deleted eigenvalues of S. If the eigenvalues are small, the residuals in the error matrix $E = S - (\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi})$ are small and the fit is good.

3.6 Bartlett's Test of Sphericity

This test is used to test the hypothesis that the correlation matrix is an identity matrix *(all the diagonal terms are 1 and all off-diagonal terms are zero).* The *P-value* is desired to be less than 0.05 since the variables to be used in Factor analysis have to be correlated. Otherwise, the variables cannot have shared variance and Factor Analysis may not be an appropriate technique for analysis of that data. what is of interest when contemplating the Factor Analysis of a correlation matrix is to determine first whether the hypothesis that all the correlations, tested simultaneously, are not statistically different from 0 can be rejected. Such a test, referred to as Barlett's test of sphericity, This is calculated as

$$\chi^2 = -\left([N-1] - \left[\frac{2k+5}{6}\right]\right) ln|R| \tag{3.32}$$

where

N is the sample size, k is the number of variables and $|R|$ is the determinant of the correlation matrix.

The degrees of freedom associated with this test statistic is $[k(k-1)]/2$, that is, the number of correlations above, or below, the main diagonal of the correlation matrix.

The Barlett's sphericity test is affected by the sample size. When N, the sample is large, as should be in Factor Analysis, the null hypothesis will almost be rejected. This is why the application of Barlett's sphericity test should be used as a lower bound to the quality of the matrix. That is, when the hypothesis that the correlation matrix is an identity matrix cannot be rejected, hence the matrix should not be factor analysed. However, rejection of the null hypothesis should not be considered as evidence that the correlation matrix is appropriate for Factor Analysis.(Pedhazur and Schmelkin, 1991)

3.7 Kaiser-Meyer-Olkin Test (KMO)

Another measure of the degree or strength of the relationship among variables to be used in Factor Analysis is the partial correlation coefficient. The partial correlations constitute the correlations between each pair of variables when the linear effects of all other variables is taken away (Norman and Streiner 2007). The *Kaiser-Meyer-Olkin Test* is a statistical technique which measures sampling adequacy, compares the magnitudes of the calculated correlation coefficients to the magnitudes of the partial correlation coefficients. This is calculated using correlations and partial correlations to test whether the variables in our sample are adequate. That is, it calculates whether variables are so highly correlated that we cannot distinguish between them (multicollinearity) and takes the following form:

$$KMO = \frac{\sum(correlations)^2}{\sum(correlations)^2 + \sum(partial\ \ correlations)^2} \tag{3.33}$$

If the variables share common factors, then it would be reasonable to expect that the partial correlation coefficients between the pairs of variables would be small when the linear effects of the other variables have been removed. The exact formula for the KMO test is given as follows:

$$\frac{\sum\limits_{i \neq j}\sum r_{ij}^2}{\sum\limits_{i \neq j}\sum r_{ij}^2 + \sum\limits_{i \neq j}\sum a_{ij}^2} \tag{3.34}$$

Where

$\sum\sum$ = sum over variables in the matrix when $i \neq j$

r_{ij} = Pearson correlation between variables i and j

a_{ij} = partial correlation coefficient between variables i and j

As can be seen from the formula for the KMO, if the sum of the squared partial correlation coefficients, a_{ij}^2, is small compared with the sum of the squared correlation coefficients, r_{ij}^2, the numerator and denominator for this test are similar and the KMO measure approaches 1. The KMO measure can range between 0 and 1, with smaller values indicating that r_{ij}^2 is small relative to a_{ij}^2 and therefore a factor analysis may be unwise (Pett et al. 2003).

When evaluating the size of the overall KMO, Kaiser, 1974, suggest using the following criteria for these values:

- Above 0.90 is "marvelous"
- In the 0.80's is "meritorious"
- In the 0.70's is just "middling"

- Less than 0.6 is "mediocre" or "unacceptable"

3.8 Factor extraction: Eigenvalues and the Scree plot

In factor Analysis procedure, not all factors are retained in the final report of the analysis. Therefore the criterion to use when deciding whether a factor is statistically important or not is very critical. Mention has to be made that eigenvalues associated with a variable indicate the substantive importance of that factor; so the eigenvalues are the λ_{ij} in equation 3.1 . Hence, it would be logical to retain only factors with large eigenvalues. Another problem to contend with is whether or not an eigenvalue is large enough to represent a meaningful factor?

Cattell, 1966, proposed graph of each eigenvalue (Y-axis) against the factor with which it is associated (X-axis). This graph is known as a scree plot because it looks like a sloping mass of loose rocks at the base of a cliff. By graphing the eigenvalues on a scree plot, the comparative significance of each factor becomes evident. Normally there will be a few factors with quite high eigenvalues, and the remaining other factors with relatively low eigenvalues. This gives the graph a very characteristic shape: there is a sharp descent in the curve followed by a tailing off, Field, 2009. It was also argued that the cut-off point for selecting factors should be at the point of inflexion of this curve. The point of inflexion is where the slope of the line changes dramatically. Thus, only factors to the left of the point of inflexion are retained (or extracted) and do not include the factor at the point of inflexion itself, Cattell, 1966.

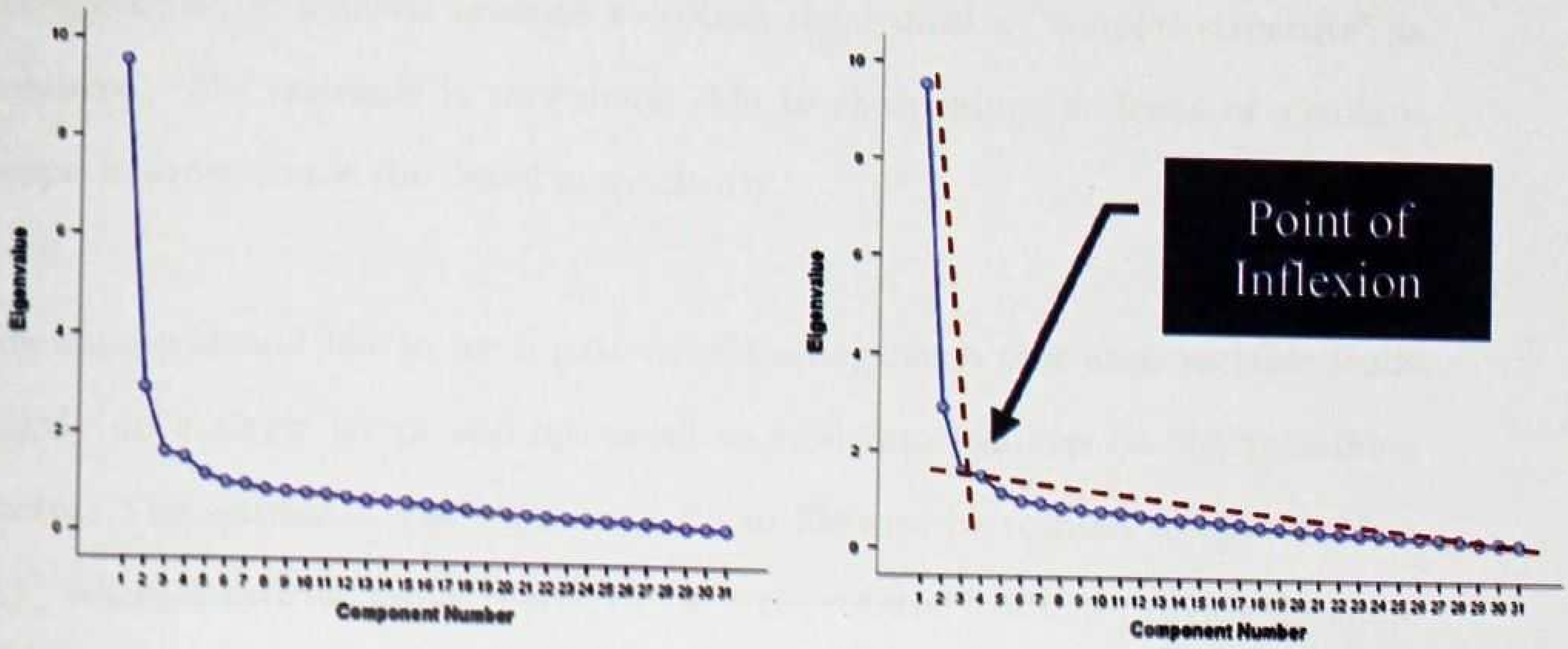

Figure 3.1: Source: A. Field, 2009. Discovering statistics using SPSS, page 640

If the sample units of data being used for the analysis is more than 200 units or cases, the scree plot provides a fairly reliable criterion for factor selection (Stevens, 2002). However, Kaiser, 1960, also advised retaining or extracting all factors with their corresponding eigenvalues greater than 1. This criterion is founded on the notion that the eigenvalues represent the amount of variation explained by a factor and that an eigenvalue of 1 represents a substantial amount of variation. In a counter argument, Jolliffe (1972, 1986) retorts that Kaiser's criterion is too strict and suggests the third option of retaining all factors with eigenvalues more than 0.7. The difference between how many factors are retained using Kaiser's methods compared to Jolliffe's can be erratic. By default, most statistical software including SPSS uses Kaiser's criterion to extract factors. Therefore, the scree plot can be used to determine how many factors are retained, then a rerun of the analysis specifying the number of factors that are required to be extracted.

3.9 Rotation

In Section 3.5, we saw that factor loadings (rows of Λ) in the general Factor model are unique only up to multiplication by an orthogonal matrix that rotates the loadings. The essence of rotation is that resultant loadings preserve the essential properties of the original loadings, that is, they reproduce the covariance matrix and satisfy all basic assumptions. Since the original loadings may not be readily interpretable, it is usual practice to rotate them until a "simpler structure" is achieved. The rationale is very much akin to sharpening the focus of a microscope in order to see the detail more clearly.

Ideally, we should like to see a pattern of loadings such that each variable loads highly on a single factor and has small to moderate loadings on the remaining factors.The estimated loading matrix $\hat{\Lambda}$ can likewise be rotated to obtain $\hat{\Lambda}^* = \Lambda T$, where T is orthogonal. Since ,$TT' = I$ the rotated loadings provide the same estimate of the covariance matrix as before.

$$S \cong \hat{\Lambda}^* \hat{\Lambda}^{*\prime} + \hat{\Psi} = \hat{\Lambda} T T' \hat{\Lambda}' + \hat{\Psi} = \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi} \tag{3.35}$$

The loadings in the i^{th} row of $\hat{\Lambda}$ comprises the coordinates of a point in the loading space corresponding to y_i geometrically. Rotation of the p points gives their coordinates with respect to new axes (factors) but otherwise leaves their basic geometric configuration intact. We hope to find a new frame of reference in which the factors are more interpretable. To this end, the goal of rotation is to place the axes close to as many points as possible. If there are clusters of points (corresponding to groupings of $y's$), we seek to move the axes in order to pass through or near these clusters. This would associate each group of variables with a factor (axis) and make interpretation more objective. The resulting axes then represent the natural factors.

If we can achieve a rotation in which every point is close to an axis, then each variable loads highly on the factor corresponding to the axis and has small loadings on the remaining factors. In this case, there is no ambiguity. Such a happy state of affairs is called simple structure, and interpretation is greatly simplified. We merely observe which variables are associated with each factor, and the factor is defined or named accordingly. In order to identify the natural groupings of variables, we seek a rotation to an interpretable pattern for the loadings, in which the variables load highly on only one factor. The number of factors on which a variable has moderate or high loadings is called the complexity of the variable. In the ideal situation referred to previously as simple structure, the variables all have a complexity of 1. In this case, the variables have been clearly clustered into groups corresponding to the factors.(Rencher and Christensen, 2012)

3.9.1 Oblique Rotation

In oblique rotation, transformation is done in a manner in which the axes do not remain perpendicular. Technically, the term oblique rotation is a misnomer, since rotation done implies that axes do not remain perpendicular. Whereas an orthogonal rotation is a transformation that preserves distances. An oblique rotation uses a general nonsingular transformation matrix Q to obtain $f^* = Q'f$, the the covariance of f^* is given as;

$$cov(f^*) = Q'\mathbb{I}Q = Q'Q \neq \mathbb{I} \tag{3.36}$$

Since distances and angles are not preserved, the communalities for f^* are different from those for f. When the axes are not required to be perpendicular, they can more easily pass through the major clusters of points in the loading space (assuming there are such clusters). Oblique axes with an angle of 38° would pass much closer to the points, and the resulting loadings would be very close to 0 and 1. However, the interpretation would not change, since the same points (variables) would be associated with the oblique axes as with the orthogonal axes. Various analytical methods for achieving oblique rotations have been proposed and are available in program packages, Direct oblimin. Typically, the output of one of these procedures includes a pattern matrix, a structure matrix, and a matrix of correlations among the oblique factors. For interpretation, we would usually prefer the pattern matrix rather than the structure matrix. The loadings in a row of the pattern matrix are the natural coordinates of the point (variable) on the oblique axes and serve as coefficients in the model relating the variable to the factors. One use for an oblique rotation is to check on the orthogonality of the factors. The orthogonality in the original factors is imposed by the model and maintained by an orthogonal rotation. If an oblique rotation produces a correlation matrix that is nearly diagonal, we can be more confident that the factors are indeed orthogonal. (Rencher and Christensen, 2012)

Chapter 4

Data Analyses and Results

4.1 Introduction

The 2008 Demographic and Health Survey organised in Ghana (2008 GDHS) is the fifth of such population and health surveys conducted in Ghana. This forms part of the global Demographic and Health Surveys (DHS) programme. The survey is organized to allow for information to monitor the population and health status in Ghana as a follow-on to the 1988, 1993, 1998 and 2003 GDHS surveys. The survey was based on a two stage sample design which uses the 2000 Population and Housing Census to produce separate estimates for key indicators for each of the ten regions in Ghana.

The GDHS 2008 household sample of more than 12,000 households was adequate to provide a sampling frame for conducting case-specific child mortality surveillance for children under five years using a Verbal Autopsy Questionnaire. Each household selected for the GDHS was eligible for interview with the Household Questionnaire, and a total of 11,778 households were interviewed. In half of the households selected for the survey, all eligible women age 15-49 and all eligible men age 15-59 were interviewed with the Women's and Men's Questionnaires, respectively. A total of 4,916 women age 15-49 and 4,568 men age 15-59 from 6,141 households were interviewed. Data collection took place over a three-month period, from early September to late November 2008. The survey obtained detailed information on fertility, marriage, sexual activity, fertility preferences, awareness and use of family planning methods, breastfeeding practices, nutritional status of women and young children, childhood mortality, maternal and child health, awareness and behaviour regarding HIV/AIDS, and other sexually transmitted infections (STI's). In addition, the 2008 GDHS collected information on domestic violence, malaria and use of mosquito nets, and carried out anaemia testing and anthropometric measurements for women and children.

The 2008 GDHS was implemented by the Ghana Statistical Service (GSS) in collaboration with the Ghana Health Service. Technical assistance was provided by ICF Macro through the MEASURE DHS programme. Financial support for the survey was provided by the U.S. Agency for International Development (USAID), the Government of Ghana, UNICEF, UNFPA and Danish International Development Agency (DANIDA). (GSS et al. 2009)

The variables used in this analysis are as follows:

1. Region
2. Type of place of residence
3. Years lived in place of residence
4. Source of drinking water
5. Religion
6. Ethnicity
7. Literacy
8. Wealth index

9. Duration of breastfeeding
10. Size of child at birth
11. When child put to breast
12. Covered by health insurance
13. Sex of child
14. Respondent's Occupation
15. Current age of child
16. Child's weight in kilograms (1 decimal)
17. Child's height in centimeters (1 decimal)
18. Haemoglobin level (g/dl - 1 decimal)
19. BMI standard deviation (new WHO)
20. Anaemia level
21. Highest educational level attended
22. Birth weight in kilograms (3 decimals)

4.2 Empirical Analysis

Initial empirical analysis was carried out on quantitative variables and categorical variables respectively. For quantitative variables, means, standard deviation and the total number of cases ~~used in~~ this analysis were calculated for each variable and the results are shown in Table 4.1. An initial sample number of 2992 was first used for the analysis. However, due the large number of missing cases recorded in almost all of the variables, SPSS option of 'Exclude cases listwise',

Table 4.1: Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Years lived in place of residence	37.5	42.5	930
Wealth index	3.0	1.3	930
Duration of breast feeding	52.4	37.6	930
Size of child at birth	2.5	1.1	930
When child put to breast	77.1	71.3	930
Current age of child	1.7	1.3	930
Child's weight in kilograms (1 decimal)	111.9	30.2	930
Child's height in centimeters (1 decimal)	827.3	120.3	930
Haemoglobin level (g/dl - 1 decimal)	96.0	17.6	930
BMI standard deviation (new WHO)	534.1	2285.0	930
Anaemia level	2.6	0.9	930
Highest educational level attended	1.8	0.7	930
Birth weight in kilograms (3 decimals)	6358.6	3431.1	930

was used in order to arrive at the 930 sample units per variable, which were used in the analysis. Since Factor analysis is being performed on the correlations (as opposed to the covariances), it is not much of an interest that variables have very different means and/or standard deviations (which is often the case when variables are measured on different scales).

Table 4.2: Type of Place of Residence

	Frequency	Percent
Urban	1000	33.4
Rural	1992	66.6
Total	2992	100.0

For the remaining 9 categorical variables, frequency distribution tables were created for each one of these variables. A total of 1000 children sampled in this Data live in urban areas, which represents 33.4% of the Type of residence. Also, a total of 1992 children, representing 66.6% children sampled in this Data live in the rural areas of the country as shown in Table 4.2. The number of children under-five who are covered by health insurance are 1202 representing 40.2%. However, 1785 children, 59.7% , are not covered by health insurance as shown in Table 4.3. The rest of the frequency tables for the other remaining variables used in this study are shown in Appendix A.

Table 4.3: Covered by Health Insurance

	Frequency	Percent
No	1785	59.7
Yes	1202	40.2
Missing	5	0.2
Total	2992	100.0

4.3 Appropriateness of the Data

In section 3.6, the Bartlett's Test of Sphericity was introduced which tests the null hypothesis that the original correlation matrix is an identity matrix. For factor analysis to work we need some relationships between variables and if the R-matrix were an identity matrix then all correlation coefficients would be zero. Therefore, we want this test to be significant (i.e. have a significance value less than .05). A significant test tells us that the R-matrix is not an identity matrix; therefore, there are some relationships between the variables we hope to include in the analysis. For this data, Bartlett's test is highly significant (p =0.00), and therefore factor analysis is appropriate as seen in Table 4.4

Also in section 3.7, the Kaiser-Meyer-Olkin Test (KMO) test was introduced and this measure varies between 0 and 1, and values closer to 1 are better. KMO measure of sampling adequacy is calculated using correlations and partial correlations to test whether the variables in our sample are adequate to correlate. For this data the value is 0.707 (Table 4.4) which falls into the range of being "middling", (Kaiser, 1974) so we should be confident that the sample size is adequate for factor analysis.

Table 4.4: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.707
Bartlett's Test of Sphericity	Approx. Chi-Square	6300.304
	df	231
	sig.	0.000

4.4 Communalities

Table 4.3 shows the table of communalities before and after extraction where communality is the proportion of common variance within a variable. Principal component method of Factor Extraction works on the initial assumption that all variance is common; therefore, before extraction the communalities are all 1 (see column labelled Initial). In effect, all of the variance associated with a variable is assumed to be common variance. Once factors have been extracted, we have a better idea of how much variance is, in reality, common. The communalities in the column labelled Extraction reflect this common variance.

We can say that 53.6% of the variance associated with 'Region' is common, or shared, variance. Also 46.5% of the variance associated with 'Type of place of residence' is common, or shared, variance. Another way to look at these communalities is in terms of the proportion of variance explained by the underlying factors. Before extraction, there are as many factors as there are variables, so all variance is explained by the factors and communalities are all 1. However, after extraction some of the factors are discarded and so some information is lost. The retained factors cannot explain all of the variance present in the data, but they can explain some. The amount of variance in each variable that can be explained by the retained factors is represented by the communalities after extraction.

Table 4.5: Communalities

	Initial	Extraction
Region	1.000	.536
Type of place of residence	1.000	.465
Years lived in place of residence	1.000	.115
Source of drinking water	1.000	.221
Religion	1.000	.020
Ethnicity	1.000	.039
Literacy	1.000	.446
Wealth index	1.000	.674
Respondent's occupation	1.000	.160
Duration of breastfeeding	1.000	.652
Size of child at birth	1.000	.065
When child put to breast	1.000	.042
Covered by health insurance	1.000	.440
Sex of child	1.000	.035
Current age of child	1.000	.855
Child's weight in kilograms (1 decimal)	1.000	.778
Child's height in centimeters (1 decimal)	1.000	.821
Haemoglobin level (g/dl - 1 decimal)	1.000	.905
BMI standard deviation (new WHO)	1.000	.295
Anaemia level	1.000	.906
Highest educational level attended	1.000	.552
Birth weight in kilograms (3 decimals)	1.000	.339

4.5 Scree Plot

The scree plot graphs the eigenvalue against the factor number. The scree plot is shown in Figure 4.1. This curve is difficult to interpret because it begins to tail off after three factors, but there is another drop after five factors before a stable flow to the graph is reached. Therefore, we could probably justify retaining either two or four factors. However, there is a fall in the graph to the fifth component number. Given the large sample, it is probably safe to assume Kaiser's criterion; however, a rerun the analysis specifying that SPSS extract only four factors is carried out.

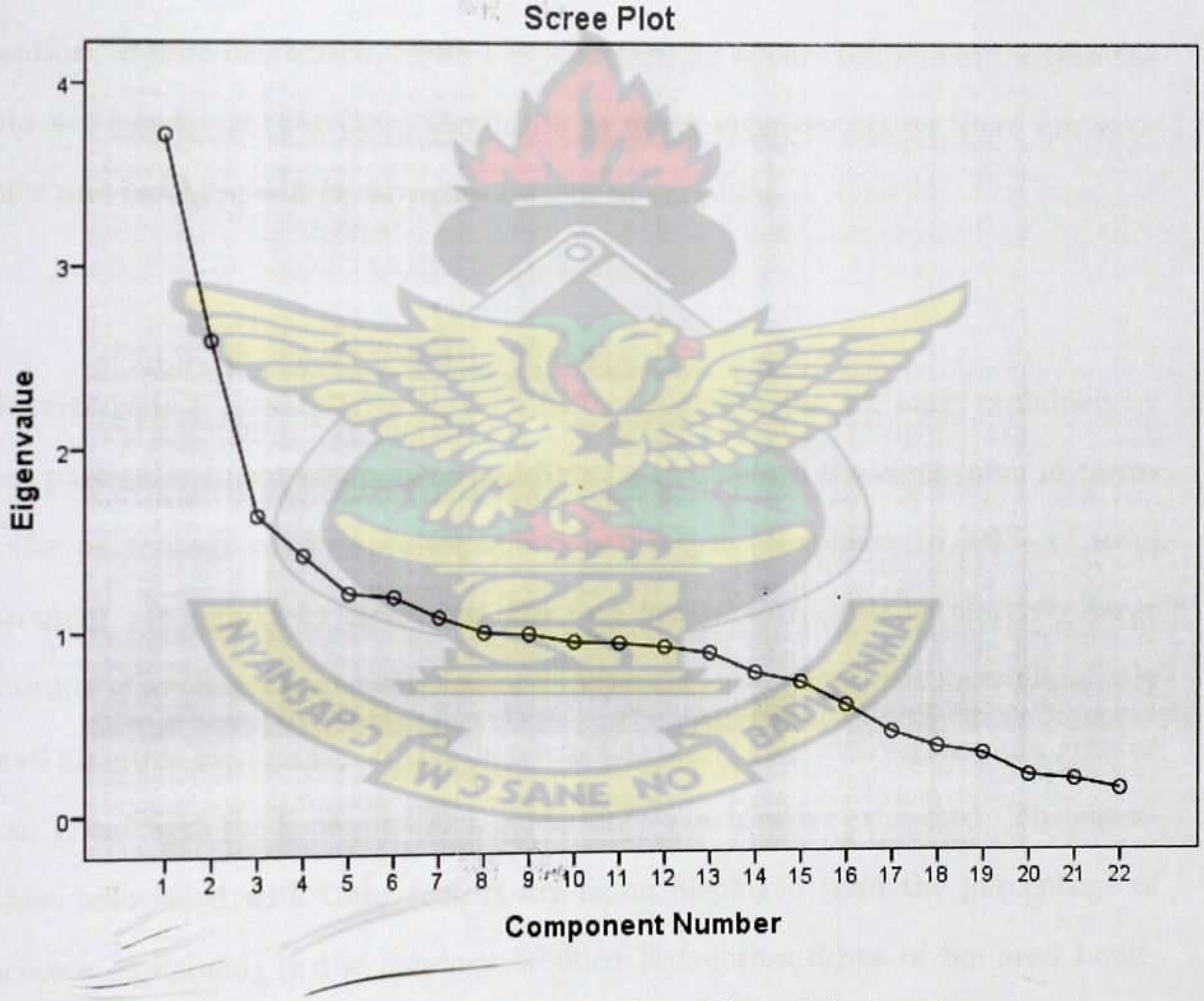

Figure 4.1: Scree plot showing the number of latent factors to extract

4.6 Factor Extraction

The first part of the factor extraction process is to determine the linear components within the data set (the eigenvectors) by calculating the eigenvalues of the R-matrix. We know that there are as many components or Factors (eigenvectors) in the R-matrix as there are variables, but most will be unimportant. To determine the importance of a particular vector we look at the magnitude of the associated eigenvalue. We can then apply criteria to determine which factors to retain and which to discard. By default SPSS uses Kaiser's criterion of retaining factors with eigenvalues greater than 1. Table 4.6 lists the eigenvalues associated with each linear component (factor) before extraction, after extraction and after rotation. Before extraction, SPSS has identified 22 linear components within the data set (we know that there should be as many eigenvectors as there are variables and so there will be as many factors as variables.

The eigenvalues associated with each factor represent the variance explained by that particular linear component and SPSS also displays the eigenvalue in terms of the percentage of variance explained (so, factor 1 explains 16.895% of total variance). It should be clear that the first few factors explain relatively large amounts of variance (especially factor 1) whereas subsequent factors explain only small amounts of variance. SPSS then extracts all factors with eigenvalues greater than 1, but with the Scree plot as a guide, only 4 factors are extracted. The eigenvalues associated with these factors are again displayed (and the percentage of variance explained) in the columns labelled Extraction Sums of Squared Loadings. The values in this part of the table are the same as the values before extraction, except that the values for the discarded factors are ignored (hence, the table is blank after the fourth factor). In the final part of the table (labelled Rotation Sums of Squared Loadings), the eigenvalues of the factors after rotation are displayed. Rotation has the effect of optimizing the factor structure and one consequence for these data is that the relative importance of the four factors is equalized. Before rotation, factor 1 accounted for considerably more variance than the remaining three (14.325%)

However for the second, third, and fourth factors, the total variance explained before rotation, 11.783%, 7.434% and 6.439% respectively, has increased to 12.377%, 9.039% and 6.810% respectively, after rotation. This is so because of the general aim of rotation which places factors as close to the axes as possible to aid in interpretation.

Table 4.6: Total Variance Extracted

Initial Eigenvalues			Extracted Sums of Squared Loadings			Rotated Sums of Squared Loadings		
Factor	Total	% of variance	Total	% of variance	cumulative %	Total	% of variance	cumulative %
1	3.717	16.895	3.717	16.895	16.603	3.151	14.325	14.325
2	2.592	11.783	2.592	11.783	28.687	2.723	12.377	26.702
3	1.635	7.434	1.653	7.434	36.112	1.989	9.039	35.741
4	1.417	6.439	1.417	6.439	42.551	1.498	6.810	42.551
5	1.207	5.488						
6	1.186	5.389						
7	1.073	4.877						
8	.987	4.487						
9	.976	4.437						
10	.932	4.235						
11	.924	4.200						
12	.901	4.097						
13	.866	3.938						
14	.756	3.438						
15	.705	3.203						
16	.578	2.628						
17	.429	1.950						
18	.347	1.578						
19	.310	1.409						
20	.187	.848						
21	.164	.746						
22	.110	.502						

4.7 Factor Matrix

This table contains the unrotated factor loadings, which are the correlations between the variable and the factor. Because these are correlations, possible values range from -1 to +1. By default SPSS displays all loadings; however, we requested that all loadings less than 0.3 be suppressed in the output and so there are blank spaces for many of the loadings.This makes the output easier to read by removing the clutter of low correlations that are probably not meaningful anyway. This matrix component matrix shows loadings of each variable onto each factor before rotation. It is particularly important for interpretation, but it is interesting to note that before rotation most variables load highly onto the first factor (that is why this factor accounts for most of the variance in Table 4.7).

Table 4.7: Factor Matrix

Factors	1	2	3	4
Region				.633
Type of place of residence	-.479	.402		
Years lived in place of residence				
Source of drinking water			.392	
Religion				
Ethnicity				
Literacy	.405	-.509		
Wealth index	.596	-.505		
Respondent's occupation		.318		
Duration of breastfeeding	-.674	-.435		
Size of child at birth				
When child put to breast				
Covered by health insurance				.530
Sex of child				
Current age of child	.711	.585		
Child's weight in kilograms (1 decimal)	.724	.487		
Child's height in centimeters (1 decimal)	.704	.541		
Haemoglobin level (g/dl - 1 decimal)	.537		.635	.406
BMI standard deviation (new WHO)			.377	-.356
Anaemia level	.514		.657	.391
Highest educational level attended	.427	-.560		
Birth weight in kilograms (3 decimals)	-.345	.408		

4.8 Rotated Factor Matrix

When an oblique rotation is conducted the factor matrix is split into two matrices: the pattern matrix and the structure matrix. The pattern matrix contains the factor loadings and is comparable to the factor matrix under orthogonal rotation.(Table 4.8) The structure matrix (Table 4.9), takes into account the relationship between factors (in fact it is a product of the pattern matrix and the matrix containing the correlation coefficients between factors). Most researchers interpret the pattern matrix, because it is usually simpler; however, there are situations in which values in the pattern matrix are suppressed because of relationships between the factors. Therefore, the structure matrix (Table 4.9) is a useful double-check and Graham et al. 2003, recommend reporting both .

The pattern matrix table contains the rotated factor loadings which represent both how the variables are weighted for each factor but also the correlation between the variables and the factor. Because these are correlations, possible values range from -1 to +1. This makes the output easier to read by removing the clutter of low correlations that are probably not meaningful anyway. However, the rotation of the factor structure has clarified things considerably and with the suppression of loadings less than 0.3 and ordering variables by loading size also make interpretation considerably easier (because there will not be the need to scan the matrix to identify substantive loadings). The next step is to look at the content of Variables that load onto the same factor to try to identify common themes. If the mathematical factor produced by the analysis represents some real-world construct then common themes among highly loading questions can help us identify what the construct might be.

The variables that load highly on factor 1 seem to all relate to Breastfeeding and growth of child. Therefore we might label this factor 'child growth per breastfeeding'. The variables that load highly on factor 2 all seem to relate to aspects of wealth and education of caregivers and parents of children under five; therefore, we might label this factor 'Wealth' . The variables which load highly on factor 3 relate to Anemia levels and Haemoglobin levels of children under five ; therefore, we might label this factor 'Vitamin A level'. Finally, variables that load highly on factor 4 all have to do with the physical location of the child; therefore, we might label this factor 'Social factor'. (Table 4.8)

Table 4.8: Pattern Matrix of Extracted Factors

Factors	1	2	3	4
Region				.716
Type of place of residence		.647		
Years lived in place of residence				
Source of drinking water				-.413
Religion				
Ethnicity				
Literacy		-.634		
Wealth index		-.770		
Respondent's occupation		.398		
Duration of breastfeeding	-.794			
Size of child at birth				
When child put to breast				
Covered by health insurance				.603
Sex of child				
Current age of child	.926			
Child's weight in kilograms (1 decimal)	.859			
Child's height in centimeters (1 decimal)	.903			
Haemoglobin level (g/dl - 1 decimal)			.946	
BMI standard deviation (new WHO)				-.491
Anaemia level			.948	
Highest educational level attended		-.708		
Birth weight in kilograms (3 decimals)		.580		

Table 4.9: Structure Matrix of Extracted Factors

Factors	1	2	3	4
Region				.712
Type of place of residence		.655		
Years lived in place of residence				
Source of drinking water				-.432
Religion				
Ethnicity				
Literacy		-.649		
Wealth index		-.787		
Respondent's occupation		.398		
Duration of breastfeeding	-.793			
Size of child at birth				
When child put to breast				
Covered by health insurance				.583
Sex of child				
Current age of child	.917			
Child's weight in kilograms (1 decimal)	.860			
Child's height in centimeters (1 decimal)	.894			
Haemoglobin level (g/dl - 1 decimal)			.936	
BMI standard deviation (new WHO)				-.498
Anaemia level			.942	
Highest educational level attended		-.722		
Birth weight in kilograms (3 decimals)		.582		

4.8.1 Relationship between Factors

Table 4.10: Component Correlation Matrix

	Child growth per breastfeeding	Wealth	Vitamin-A Levels	Social Factors
Child growth per breastfeeding	1.000	-.052	-.064	-.029
Wealth	-.052	1.000	-.110	-.009
Vitamin-A Levels	-.064	-.110	1.000	-.121
Social Factors	-.029	-.009	-.121	1.000

The final part of the output is a correlation matrix between the factors (Table 4.10). This matrix contains the correlation coefficients between factors. From Table 4.10, it is observed that the correlation coefficient between 'Child growth per breast feeding' and 'Wealth' is -0.052 which is a negatively weak correlation. This implies that although there is not much relationship between these latent factors, an increase in the wealth status of parents or care givers of children under-5 have detrimental effects on the nutritional status of children. This can be attributed to the fact that parents, especially mothers would not have enough time at home to breastfeed the child. This may lead to the child having low scores on his or her weight-for-age, height-for-age, and weight-for-height Standard Deviation scale and can lead to the child being stunted, wasted and/or underweight.

The correlation coefficient between 'Child growth per breast feeding' and 'Vitamin-A level' is also -0.064 (Table 4.10). This represents a negatively weak relationship between these latent factors. It can hence be argued that if a child is weaned off breast-milk before reaching at least 18 months, and given solid food, the child may be malnourished. This is so because breast-milk contain much needed nutrients to aid in child growth and development. Thus is the child is given food especially, those rich in Vitamin-A, child growth can be hampered causing stunted growth in the child. It is however, advised that children are gradually weaned off breast-milk and introduced to solid foods rich in Vitamin A and other essential Vitamins after the child is 18 months old.

'Child growth per breast feeding' and 'Social Factors' also have a negatively weak correlation coefficient ,-0.029 (Table 4.10). This is mainly due to the rapid increase in urbanisation and expansion of social amenities. Though improved social factors is good as access to food, health facilities and inoculations against childhood diseases are greatly increased. However, not all change is desirable, and not all new food habits are good especially for children under-5. There are harmful effects of the rapid spread of bottle-feeding using infant formula or animal milk in place of breastfeeding. This is an undesirable, relatively new food trend for children. Locally available complementary or weaning foods, home-produced and traditionally fed, are often as or more nutritious than the manufactured baby foods, and then are always much cheaper.(FAO, 2013)

In general, the latent factors here have little or no relationship with any other factors (correlation coefficients are low). This is an indication of the separate nature of the variables that engendered these factors. Hence any policy aimed at addressing the problem of malnutrition should be targeted at these factors individually.

Chapter 5

Summary,Conclusion and Recommendations

5.0 Introduction

This Chapter five presents synergy of this thesis, summary of findings from the study and gives a conclusion statement based on the findings. It also suggests further research areas for other researchers.

5.1 Summary of Findings from the Study

Factor Analysis which has a main objective of finding latent factors among numerous variables was used in this study. The objective of this study was then to use Factor Analysis to find the latent factors Associated with the determinants of malnutrition in Ghana using data on Under-Five children from the 2008 Ghana Demographic and Health Survey, from which 22 variables were employed for the analysis.

To ascertain the appropriateness of this data, two critical tests were carried out. The first was the Kaiser-Meyer-Olkin measure of Sampling Adequacy which gave a test statistic of 0.707 indicating the appropriateness of the sample size for Factor

Analysis. Next, the Bartlett's Test of Sphericity, a measures of singularity of the matrix of correlations among these 22 variables, gave a p-value=0.000 indicating that the correlation matrix used for this analysis was not an identity matrix and that the variables had significant association among themselves.

The Principal Component method of Latent Factor extraction was used with the aid of a Scree plot to obtain four latent factors. (Figure 4.1 and Table 4.6). The Rotated Factor Matrix of Extracted Factors (Table 4.8) gave a clear idea as to the nature of these four latent factors. Parsimony was then achieved in this analysis since four latent factors were obtained in this study after the whole procedure for Factor Analysis was carried out.

The four Latent factors which are meant for predictive purpose of malnutrition of under-5 children are:

1. Child growth per rate of Breast feeding

2. Wealth

3. Vitamin-A level

4. Social factor

A summary of how these latent factors are associated with the variables used in this study is given below;

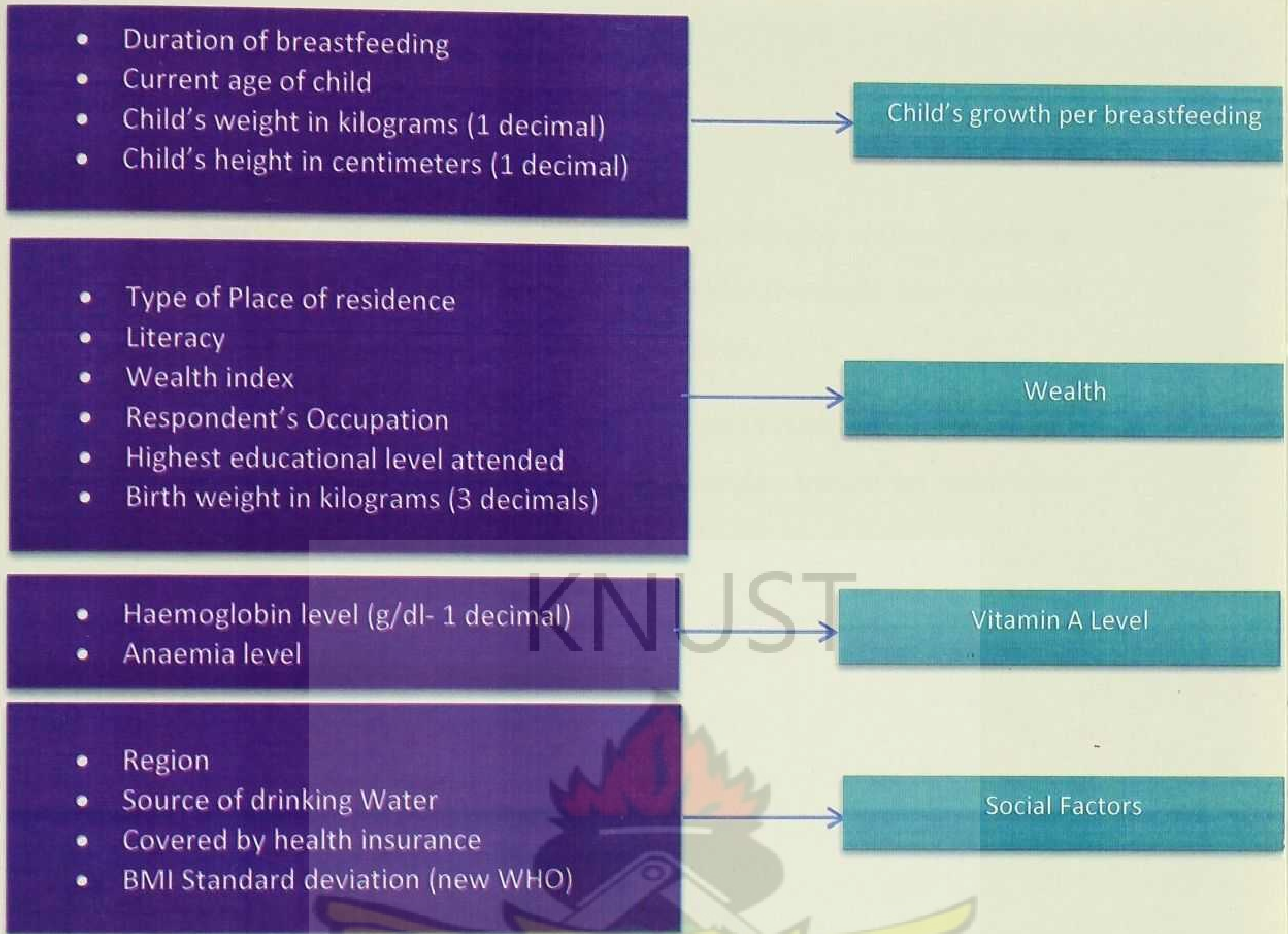

Figure 5.1: Diagram showing how Factors are associated with variables

The Latent factors obtained in this study, (Child growth per breastfeeding,Wealth, Vitamin-A Level and social factors) are in line with similar ones UNICEF have identifed as among the main causal factors of child malnutrition. According to UNICEF, 2013, low status of women, poor maternal nutrition, inadequate pre-natal care and the disproportionate burden of physical labour borne by mothers are some of the greatest impediments to improving nutritional status of children.

5.2 Conclusion

In Section 4.8.1, the association between these latent factors was determined. It was found out that there was weak correlations among these four latent factors. Hence in conclusion, policies and efforts to help address the problem if Malnutrition among children under-5 should be aimed at addressing these latent factors individually and not in a congregate manner.

Recommendations

1. A further study could be carried out to find how each of these four factors affect the Body Mass Index directly which is the single most significant index to measure nutritional status in children.

2. An extension of this study can be carried out to other age brackets, that is adult population to find out if these factors also account for malnutrition in the entire population.

References

Adelman, Irma, and Morris, C. T. (1965). Factor analysis of the interrelationship between social and political variables and per capita gross national product. *Quarterly Journal of Economics*, *79*, 555-578.

Advameg. (2013, February). *ma.* Available from http://www.faqs.org

Agnarsson, I., Mpell, G., Gunnlaaugsson, Y., Hofuander, T., and Greider. (2001). Infant feeding practices during the first months of life in a rural area,tanzania. *East Africa Medical Journal*, *78 (1)*, 9-13.

Appoh, L. Y., and Krekling, S. (2005). Maternal nutritional knowledge and child nutritional status in the volta region of ghana. *Maternal & child nutrition*, *1(2)*, 100-110.

Bhavsar, S., Hemant, M., and Kulkarni, R. (2012). Maternal and environmental factors affecting the nutritional status of children in mumbai urban slum. *International Journal of Scientific and Research Publications*, *2(11)*.

Bhutta, Z. A., Ahmed, T., Black, R. E., Cousens, S., Dewey, K., Giugliani, E., et al. (2008). What works? interventions for maternal and child undernutrition and survival. *The Lancet*, *371 (9610)*, 61693-6.

Black, R. E., Cousens, S., Johnson, H. L., Lawn, J. E., Rudan, I., Bassani, D. G., et al. (2010). Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet*, *375(9730)*, 1969-1987.

Cattell, R. B. (1966a). The meaning and strategic use of factor analysis. *Handbook of Multivariate Experimental Psychology. Chicago: Rand McNally*, 174-243.

Cattell, R. B. (1966b). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245276.

Cheah, W. L., Manan, W. M. W. A., Zabidi-Hussin, Z., and Chang, C. T. (2009). Development of a questionnaire for the study of malnutrition among children in rural kelantan. *Malaysia Rural and Remote Health*, *9*, 1155.

Chinebuah-Aidam, B. A., Perez-Escamilla, R., and Lartey, A. (2004). A randomized trial on the impact of peer counseling on exclusive breast-feeding in ghana: controlling for the hawthorne effect. *FASEB J*, *18*.

CIA. (2013, January). *Ghana.* Available from `https://www.cia.gov`

Darlington, R. B., Weinberg, S., and (1973), H. W. (1973). Canonical variate analysis and related techniques. *Review of Educational Research*, 453-454.

Dewey, K. G., Cohen, R., Brown, K., and L. I. Vera, L. (1999). Age of introduction of complementary foods and growth and low-birth-weight, breastfed infants. *American Journal of Clinical Nutrition*, *1(2)*, 32-45.

Duggan, C., Watkins, J. B., and Walker, W. A. (2008). *Nutrition in pediatrics: Basic science, clinical applications* (W. A. W. John B. Watkins, Ed.). PMPH-USA.

FAO. (2013, April). *Human nutrition in the developing world.* Available from `http://www.fao.org`

Field, A. (2009). *Discovering statistics using spss.* . Sage Publications Limited.

Flood, A., Rastogi, T., Wirflt, E., Mitrou, P. N., Reedy, J., Subar, A. F., et al. (2008). Dietary patterns as identified by factor analysis and colorectal cancer among middle-aged americans. *The American journal of clinical nutrition*, *88(1)*,, 176-184.

GhanaDistricts. (2013, January). *Ghana districts.* Available from `http://www.ghanadistricts.com/home/`

Ghartey, A. B. (2010). Nutrition policy and programs in ghana: The limitation of a single sector approach. *The International Bank for Reconstruction and Development / The World Bank.*

GHS. (2007). Annual report.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, *68(3)*, 532-560.

(GSS), G. S. S., (GHS), G. H. S., and Macro, I. (2009). 2008 demographic and health survey, ghana report.

Gupta, A., Mathur, G. P., and Sobti, J. C. (2002). World health assembly recommends exclusive breastfeeding for first six months. *Journal of the Indian Medical Association*, *100(8)*, 510- 515.

Hien, N. N., and Hoa, N. N. (2009). Nutritional status and determinants of malnutrition in children under three years of age in nghean, vietnam. *Pakistan Journal of Nutrition*, *8(7)*, 958-964.

Independent, T. (2007, January). *Birth rates must be curbed to win war on global poverty.* Available from ind

Johnson, R. A., and Wichern, D. W. (2002). *Applied multivariate statistical analysis (vol. 5, p. 767).* Prentice hall Upper Saddle River, NJ.

Jolliffe, I. T. (1972). Discarding variables in a principal component analysis. i: Artificial data. *Applied statistics*, 160-173.

Jolliffe, I. T. (1986). *Principal components in regression analysis. in principal component analysis.* Springer New York.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement.*, *20*, 141-151.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39(1)*, 31-36.

Lioret, S., Touvier, M., Lafay, L., Volatier, J. L., and Maire, B. (2008). Dietary and physical activity patterns in french children are related to overweight and socioeconomic status. *The Journal of nutrition*, *138(1)*, 101-107.

McArdle, J. J. (1990). Principles versus principals of structural factor-analyses. *Multivariate Behavioral Research*, *25(1)*, 81-87.

McKeever, P., and Miller, K. L. (2004). Mothering children who have disabilities:

A bourdieusian interpretation of maternal practices. *Social Science and Medicine*, *59*, 11771191.

MedicineNet. (2013, February). Available from `http://www.medterms.com`

MedicineOnline. (2013, February). *Malnutrition.* Available from `http://www.medicineonline.com`

MOFA. (2013, March). *food-security-situation-in-ghana.* Available from `http://mofafoodsecurity.wordpress.com`

MotherChildnutrition. (2013, March). *Impact of malnutrition.* Available from `http://motherchildnutrition.org`

Nemer, L., Gelband, H., and Jha, P. (2001). The evidence base for interventions to reduce malnutrition in children under five and school-age children in low and middle-income countries. *Commission on Macro-economics and Health, CMH Working Paper Series*, *5(11)*.

Newby, P. K., Muller, D., Hallfrisch, J., Andres, R., and Tucker, K. L. (2004). Food patterns measured by factor analysis and anthropometric changes in adults. *The American journal of clinical nutrition*, *80(2)*, 504-513.

Norman, G. R., and Streiner, D. L. (2007). *Biostatistics: the bare essentials.* . Pmph-usa.

Ojofeitimi, E. O., Owolabi, O. O., Aderonmu, A., Esimai, A. O., and Olasanmi, S. O. H. (2003). A study on under nutritional status and its determinants in a semi-rural community of ile-ife,osun state,nigeria,nutri.health. *Nutrition and Health,*, *(17)1*, 21-27.

Onis, M. de, Blssner, M., Borghi, E., and Frongillo, E. A. (2004). Meeting international goals in child malnutrition. *Journal of the American Medical Association*, *291*, 26002606.

Oocities. (2013, January). *Nutrition.* Available from `http://www.oocities.org/cspslibrary/nutrition`

Pedhazur, E. J., and Schmelkin, L. P. (1991). . *measurement, design, and analysis: An integrated approach.* . Psychology Press.

Pett, M. A., Lackey, N. R., and Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research.* SAGE Publications, Incorporated.

Rencher, A. C., and Christensen, W. F. (2012). *Methods of multivariate analysis (vol. 709).* Wiley.

Rikimaru, T., Taniguchi, K., Yartey, J. E., Kennedy, D. O., and Nkrumah, F. K. (1998). Humoral and cell-mediated immunity in malnourished children in ghana. *European journal of clinical nutrition*, *52(5)*,.344.

Ruel, M., Habicht, J., Pindtrup-Anderson, P., and Grohn, Y. (1992). The effects of maternal nutrition knowledge on the association between maternal schooling and child nutrition status in lesotho. *American Journal of Epidemiology*, *135*, 904914.

Rummel, R. J. (1967). Understanding factor analysis. *Journal of Conflict Resolution*, 444-480.

Rummel, R. J. (1970). Applied factor analysis. *Northwestern University Press.*.

Service, G. G. H. (2005). Imagine ghana free of malnutrition.

Shi, L., and Singh, D. A. (2009). Delivering health care in america. *Jones & Barlett Learning.*

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences.* Hillsdale, NJ: Erlbaum.

Thairu, L. N., Pelto, G. H., Rollins, N. C., Bland, R. M., and Ntshangase, N. (2005). Socio-cultural influences on infant feeding decisions among hiv-infected women in rural kwa-zulu natal,south africa. *Maternal and Child Nutrition*, *1(1)*, 2-10.

Thomson, G. (1947). Charles spearman. 1863-1945. *Obituary Notices of Fellows of the Royal Society*, *5(15)*, 373.

UNICEF. (2006). Progress for children: A report card on nutrition.

UNICEF. (2012, March). *Facts for life.* Available from `http://www.factsforlifeglobal.org`

Wagstaff, A., and Watanabe, N. (1999). Socioeconomic inequalities in child malnutrition in the developing world. *World Bank Policy Research Working Paper, 2434*.

WHO. (2013, February). Available from http://www.who.int

Wikipedia. (2013, February). *Factor analysis.* Available from http://en.wikipedia.org

Wikispaces. (2013, February). *Malnutrition.* Available from http://medanth.wikispaces.com

Wiley, A., and Allen, J. S. (2009). *Medical anthropology: A biocultural approach.* Oxford University Press.

Wolff, C. B., and Wolff, H. K. (1995). Maternal eating patterns and birth weight of mexican american infants. *Nutrition and health, 10(2)*, 121-134.

Worldbank. (2013, January). *Ghana.* Available from http://web.worldbank.org

Wunderlich, G. S., and Norwood, J. L. (2006). *Food insecurity and hunger in the united states: An assessment of the measure.* (D. o. B. Committee on National Statistics, S. Sciences, and D. Education. Washington, Eds.). The National Academies Press.

Appendix A

Ethnicity

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Akan	1134	37.9	37.9	37.9
	Ga/Dangme	141	4.7	4.7	42.6
	Ewe	367	12.3	12.3	54.9
	Guan	81	2.7	2.7	57.6
	Mole-Dagbani	767	25.6	25.7	83.3
	Grussi	161	5.4	5.4	88.7
	Gruma	201	6.7	6.7	95.4
	Mande	24	.8	.8	96.2
	Other	114	3.8	3.8	100.0
	Total	2990	99.9	100.0	
Missing	99	2	.1		
Total		2992	100.0		

Figure A.1: Ethnicity

Literacy

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Cannot read at all	2131	71.2	71.3	71.3
	Able to read only parts of sentence	319	10.7	10.7	82.0
	Able to read whole sentence	527	17.6	17.6	99.6
	No card with required language	9	.3	.3	99.9
	Blind/visually impaired	3	.1	.1	100.0
	Total	2989	99.9	100.0	
Missing	9	3	.1		
Total		2992	100.0		

Figure A.2: Literacy

Region

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Western	270	9.0	9.0	9.0
	Central	227	7.6	7.6	16.6
	Greater Accra	279	9.3	9.3	25.9
	Volta	245	8.2	8.2	34.1
	Eastern	261	8.7	8.7	42.8
	Ashanti	439	14.7	14.7	57.5
	Brong Ahafo	266	8.9	8.9	66.4
	Northern	479	16.0	16.0	82.4
	Upper East	227	7.6	7.6	90.0
	Upper West	299	10.0	10.0	100.0
	Total	2992	100.0	100.0	

Figure A.3: Region

Religion

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Catholic	423	14.1	14.2	14.2
	Anglican	18	.6	.6	14.8
	Methodist	141	4.7	4.7	19.5
	Presbyterian	162	5.4	5.4	24.9
	Pentecostal/Charismatic	943	31.5	31.5	56.4
	other Christian	290	9.7	9.7	66.1
	Moslem	602	20.1	20.1	86.3
	Traditional/spiritualist	256	8.6	8.6	94.8
	No religion	150	5.0	5.0	99.9
	Other	4	.1	.1	100.0
	Total	2989	99.9	100.0	
Missing	99	3	.1		
Total		2992	100.0		

Figure A.4: Religion

Sex of child

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	1526	51.0	51.0	51.0
	Female	1466	49.0	49.0	100.0
	Total	2992	100.0	100.0	

Figure A.5: Sex of Child

Source of drinking water

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Piped into dwelling	87	2.9	2.9	2.9
	Piped to yard/plot	152	5.1	5.1	8.0
	Public tap/standpipe	693	23.2	23.2	31.2
	Tube well or borehole	1224	40.9	40.9	72.1
	Protected well	131	4.4	4.4	76.5
	Unprotected well	78	2.6	2.6	79.1
	Protected spring	1	.0	.0	79.1
	Unprotected spring	37	1.2	1.2	80.3
	River/dam/lake/ponds/stream/canal/irrigation channel	393	13.1	13.1	93.5
	Rainwater	11	.4	.4	93.8
	Tanker truck	7	.2	.2	94.1
	Cart with small tank	4	.1	.1	94.2
	Bottled water	6	.2	.2	94.4
	Sachet water	143	4.8	4.8	99.2
	Not a dejure resident	24	.8	.8	100.0
	Total	2991	100.0	100.0	
Missing	99	1	.0		
Total		2992	100.0		

Figure A.6: Source of drinking water

Respondent's occupation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Unemployed/seeking employment	300	10.0	10.1	10.1
	Medical & Related Workers	7	.2	.2	10.3
	Accountants	2	.1	.1	10.4
	Teachers	42	1.4	1.4	11.8
	Professional, Technical and Related Workers Not Elsewhere Classified	2	.1	.1	11.9
	Managers (GM, MD's & Executive Directors)	3	.1	.1	12.0
	Stenographers, Typists	14	.5	.5	12.4
	Book-Keepers, Cashiers and Related Workers	4	.1	.1	12.6
	Transport Conductors	3	.1	.1	12.7
	Telephone and Telegraph Operators	2	.1	.1	12.7
	Managers (Wholesale & Retail Trade)	9	.3	.3	13.0
	Working Proprietors (Wholesale & Retail Trade)	12	.4	.4	13.4
	Technical Salesmen, Commercial Travellers and Manufacturers' Agents	1	.0	.0	13.5
	Insurance, Real Estate, Securities & Business Services Salesmen & Auctioneers	3	.1	.1	13.6
	Salesmen, Shop Assistants & Related Workers (Newspapers Stand)	9	.3	.3	13.9
	Wholesale Market Traders	13	.4	.4	14.3
	Retail Market Traders	441	14.7	14.8	29.1
	Hawkers, Street and Pavement Vendors	303	10.1	10.2	39.3
	Sales Workers not Elsewhere Classified	158	5.3	5.3	44.6
	Managers (Catering & Lodging Services)	2	.1	.1	44.7

Figure A.7: Respondent's occupation 1

Respondent's occupation

	Frequency	Percent	Valid Percent	Cumulative Percent
Working Proprietors (Catering & Lodging Services)	2	.1	.1	44.7
Housekeeping & Related Service Supervisors	1	.0	.0	44.8
Cooks, Waiters, Bartenders & Related Workers	51	1.7	1.7	46.5
Maids, Related Housekeeping Service Workers not Elsewhere Classified	1	.0	.0	46.5
Building Caretakers, Charworkers, Cleaners & Related Workers (Washmen/women)	6	.2	.2	46.7
Launderers, Dry-cleaners and Pressers	1	.0	.0	46.7
Hairdressers, Barbers, Beauticians and Related Workers	126	4.2	4.2	51.0
Protective Service Workers (Private Security, Watch dog Committee, Watchmen)	7	.2	.2	51.2
Service Workers not Elsewhere Classified	1	.0	.0	51.2
Farmers	1158	38.7	38.9	90.1
Agriculture & Animal Husbandry Workers	3	.1	.1	90.2
Fishermen, Hunters & Related Workers	29	1.0	1.0	91.2
Miners, Quarrymen, Well Drillers and Related Workers	8	.3	.3	91.5
Wood Preparation Workers and Paper Makers	4	.1	.1	91.6
Chemical Processors and Related Workers	1	.0	.0	91.6
Spinners, Weavers, Knitters, Dyes and Related Workers	17	.6	.6	92.2
Food & Beverage Processors	58	1.9	1.9	94.2

Figure A.8: Respondent's occupation 2