DISCRIMINANT ANALYSIS AS A TOOL FOR UNDER FIVE ANTHROPOMETRIC CLASSIFICATION IN GHANA

By:

Abukari Alhassan

UDS/MAS/0001/08

June 2010

DISCRIMINANT ANALYSIS AS A TOOL FOR UNDER FIVE ANTHROPOMETRIC CLASSIFICATION IN GHANA

A THESIS SUBMITTED TO THE DEPARTMEMNT OF STATISTICS,

UNIVERSITY FOR DEVELOPMENT STUDIES, GHANA

IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE

By:

Abukari Alhassan

UDS/MAS/0001/08

Dedication

With endless thoughts to my children, Farida, Zaidan and Einus for their endless love and care. And to my lovely Dad for his moral support.

Declaration

I hereby declare that this submission is my own work and to the best of my knowledge, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma at UDS or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by colleagues, with whom I have worked at UDS or elsewhere, during my candidature, is fully acknowledged.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.

………………………… Student Name	…………………. Signature	………......... Date

Certified by:

…………………… Supervisor	………………… Signature	…......... Date

Certified by:

…………………… Head of Department	……………………. Signature	………......... Date

Abstract

Under five years malnutrition remains an important public health and development problem in developing countries. In this dissertation, anthropometric indicators (height-for-age, weight-for-height and weight-for-age) of children(less than 5 years) are used as indicators for classifying the nutritional status of children. The study captured 2992 children under five in Ghana from the 2008 Ghana Demographic and Health Survey (GDHS) data set. The multivariate technique of discriminant function analysis was used to classify the nutritional status using selected variables consisting of measured attributes of children and their mothers. The classification function for the discriminant analysis classified correctly 77.9%, 92.4% and 88.4% of the group cases respectively for stunted, wasted and underweight classifications. Logistic regression was used as an investigative tool. The study further identified that the probability of a child being chronically malnourished is certain (1) if the child has low score of BMI. Results from logistic regression, showed that if repeated samples are taken, we are 95% confident that between 0.991 and 0.996 of the malnourished children will be classified. The study identified the body mass index of the child as the major determining factor in classifying the nutritional status of children under five in Ghana. On the basis of the analyses it was recommended that sufficient controls should be exercised in the discriminant function of the three nutritional categories.

Acknowledgments

First and foremost, I would like to express my sincere thanks to the ALMIGHTY GOD for the gift of writing and the wisdom bestowed on me. I owe a great debt of gratitude to my supervisor, Dr. (Mrs.) A. O. Adebanji for her time, patience and excellent guidance through constructive comments and suggestions which have made this research a success.

I would also like to acknowledge the contribution of Prof. K. S. Nokoe, for his leading and motivative role throughout my course work and the research. My sincere thanks also goes to the Dean of the Faculty of Mathematical Sciences, Dr. Oyetunji for his fatherly role.

I also thank the Head of Department for Mathematics Dr. Alhassan Seidu for giving me an insight into basic ethics in research.

My appreciation also goes to the post graduate coordinator Mr. K. A. Gbolagade for his advisory role.

I also thank Dr. Matilda Steiner - Aseidu a Senior Lecturer in the department of Nutrition and Food sciences in the University of Ghana for giving me an insight into Nutrition.

The effort of Peter Takyi Peprah of Ghana Statistical Service for extracting the data for me is greatly appreciated.

I also thank Baba Seidu of the Mathematics department of the University for Development Studies for his immense contribution during the typesetting of the work.

Finally, to all my friends, especially Ibrahim Mohammed Daabo of the Tamale Polytechnic, I say a big thank you.

Contents

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Background

Adequate nutrition is critical to child development. The period from birth to five years of age is important for optimal growth, health and development. Unfortunately, this period is often marked by growth faltering, micronutrient deficiencies and common childhood illnesses such as diarrhoea and Acute Respiratory Infections (ARI).

All over the world children less than five years have been considered to be vulnerable to disease conditions. The 2008 Demographic and Health Survey (DHS) have given prominence of the nutritional status of children under five years.

Current day usage of nutrition and health indicators to gauge the severity of famine and complex emergencies originates with the early experiences of Non-Governmental Organizations (NGO'S), during famine and complex emergencies in the late sixties in Biafra and India. By the seventies, nutritional surveys in emergencies were increasingly common.

There are statistical criteria that classify the nutritional status of individual children by comparison with a reference population, such as weight for height (WH), weight for age (WA) and height for age (HA) based on the National Centre for Health Statistics/Centre for Disease Control/World Health Organization (NCHS/CDC/WHO) population. These referee e values and associated cutoff points are a tool for analyzing data and providing a common (international) basis for comparing population.

Malnutrition classification systems have an equally long and convoluted history. The first cross classification of wasting and stunting by Waterlow (1972) was originally intended to distinguish patterns of severe malnutrition among the children admitted to hospitals, but since then quantitative classification of wasting, and stunting has been used in community studies of prevalence and severity.

Waterlow's advice of 37 years ago is still relevant to the full security (situation analysis) of children under five years today. He said "a classification system needs to be simple, and ideally a classification of protein energy and malnutrition will take account simultaneously three factors: Quality of type of growth failure, severity and duration. Howe Devereux (2004) also drew attention to magnitude as an important aspect of classification system. For the magnitude includes both the skill and density of the phenomena in terms of population numbers affected, geographic spread and density (Devereux, 2008). In summary, to be useful, anthropometric classification systems needs to be simple and take account of the type of growth failure, its severity, duration and magnitude and needs to be agreed upon by key stakeholders.

Anthropometric indicators are most commonly used as proxies for 'nutritional status' and are constructed from nutritional indices. The use of nutritional indices and indicators and their interpretation differ according to whether they relate to individuals or populations. Our major concern in this study is their use as indicators at population level, which conceptually is very different from the individual diagnoses of malnutrition and has important implications for non-nutritional variables collected.

This study presents information on discriminating the anthropometric classifications of young children under five years and key variables of the characteristics of the mother are used in building a discriminant function.

It makes use of the 2008 Ghana Demographic and Health Survey (GDHS). We would also explore the data to determine how anthropometric indicators of child nutrition reflect the health of children in Ghana.

1.2 Demographic Profile of the Study Area

There are a variety of sources that provide demographic information about the Ghanaian population, including censuses, administrative/routine data, and surveys. Population censuses provide more comprehensive demographic information than all the other sources.

Ghana has undertaken four censuses since independence in 1957. The first post-independence census was conducted in 1960, reporting a population of 6.7 million. However, because population censuses are resource intensive, thus expensive to implement, and generally take place at intervals of ten years, sample surveys are important for informing demographic profiles.

During inter-censual periods, sample surveys are conducted to collect a wide range of data to complement the census data. Because sample surveys are cheaper and can be implemented more quickly, they are conducted at regular intervals. The Ghana Demographic and Health Survey (GDHS), which is a household survey, is an example of the collection of sample survey data.

One other important but often neglected data source in Ghana is administrative (or routine) data. These data are generated as a by-product of events and processes and they provide relatively up to-date information to fill the data gaps in both censuses and surveys. Vital registration systems (birth and death registration), health systems (immunizations), and education data (enrollment) are examples of administrative data.

1.3 Population Policy and Reproductive Health Programmes

The1969 National Population Policy was revised in 1994 after 25 years of implementation. The revision took into account emerging issues such as HIV/AIDS, population and the environment, concerns about the elderly and children, and the development of new strategies to ensure achievement of the revised policy objectives. The revision of population policy also entailed concerted effort to systematically integrate population variables in all areas of development planning. According to GDHS 2008, the major goals of the revised population policy include:

- Reducing the total fertility rate from 5.5 in 1993 to 5.0 by the year 2000, 4.0 by 2010, and 3.0 by 2020. Accordingly, the policy aims at achieving a contraceptive prevalence rate (CPR) of 15 percent for use of modern methods by the year 2000, 28 percent by 2010, and 50 percent by the year 2020 (GDHS, 2008).
- Reducing the population growth rate from about 3 percent per annum to 1.5 percent by the year 2020; and
- Increasing life expectancy from the current level of 58 years, to 65 years by 2010, and to 70 years by 2020.

The attainment of these population goals is recognised as an integral component of the national strategy to accelerate economic development, eradicate poverty, and enhance the quality of life of all Ghanaians.

In collaboration with the United Nations Population Fund (UNFPA), the United States Agency for International Development (USAID), the World Bank, and other development partners, Ghana has implemented several projects aimed at reducing reproductive health problems in the population. Support from these agencies has targeted policy coordination, implementation, and service delivery.

The government is committed to improving access and equity of access to essential health care services. The priority areas identified include addressing the problems of HIV/AIDS and other sexually transmitted infections (STIs), malaria, tuberculosis, guinea worm disease, poliomyelitis, reproductive health, maternal and child health, accidents and emergencies, non-communicable diseases, oral health and eye care, and specialised services. Emphasis is also being placed on preventive as well as community-based health care services.

The scare associated with the spread of HIV/AIDS attracted considerable attention from the government and its development partners. The government set up the National AIDS Commission to oversee the implementation of HIV/AIDS programmes using a multi-sectoral approach. This was to ensure that HIV/AIDS prevention education, treatment, care and support reached every corner of the country. The Ghana Health Service (GHS) also set up the National AIDS Control Programme (NACP) to offer HIV/AIDS prevention education and services. The combined efforts of all stakeholders ensured the implementation of the Ghana HIV/AIDS Strategic Framework: 2001-2005 (World Bank, 2003). This collaborative effort had a positive impact and in 2003 only 2 percent of Ghanaian adults had contracted HIV (Ghana Statistical Service, 2004). This level is expected to decline. Roll back malaria, tuberculosis (TB-DOTS), and Integrated Management of Childhood Illnesses (IMCI) are still priority areas under the country's health care system. Other health interventions instituted as part of government's efforts to make health care accessible and affordable to all include the introduction of the National Health Insurance Scheme (NHIS) and the free maternal care programme (United Nations, 2008).

1.4 Measurement of Nutritional Status Among Children

The 2008 GDHS collected information on the nutritional status of children under five by measuring the height and weight of all children under six years of age. The measurements were collected with the aim of calculating three indices-weight-for-age, height-for-age, and weight-for-height-all of which take age and sex into consideration. Weight measurements were obtained using lightweight, electronic Seca scales with a digital screen, designed and manufactured under the guidance of the United Nations Children's Fund (UNICEF). Height measurements were carried out using a measuring board produced by Shorr Productions. Children younger than 24 months were measured lying down (recumbent length) on the board while standing height was measured for older children. 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 Multicentre Growth Reference Study (WHO, 2006).

Each of the three nutritional status indicators described below is 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.

These indices height for age, weight for height, and weight for age provides different information about growth and body composition that is used to assess nutritional status.

The height-for-age index is an indicator of linear growth retardation and cumulative growth deficits. Children whose height-for-age Z-score is below minus two standard deviations (-2 SD) are considered short for their age (stunted) and are chronically malnourished. Children who are above (-2SD) are considered normal. Stunting reflects failure to receive adequate nutrition over a long period of time and is also affected by recurrent and chronic illness. Height-for-age, therefore, represents 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 measures body mass in relation to body height or length and describes current nutritional status. Children with Z-scores below -2 SD are considered thin (wasted) and are acutely malnourished. Wasting represents the failure to receive adequate nutrition in the period immediately preceding the survey and may be the result of inadequate food intake or a recent Episode of illness causing loss of weight and the onset of malnutrition. Children whose weight-for-height is above - 2SD are considered normal.

Weight-for-age is a composite index of height-for-age and weight-for-height. It takes into Account both acute and chronic malnutrition. Children whose weight-for-age is below -2 SD are classified as underweight. Children whose weight-for-age is above (-2 SD) are considered normal.

For the purpose of our study, we focus on the three nutritional indices, thus stunting, wasting and underweight as the three populations from which we carry out our classifications based on the variables defined in chapter three.

1.5 Problem Statement

Child nutrition is a world-wide phenomenon that has attracted a lot of research over the past decades. Special emphasis is placed on the anthropometric indices (Height-for-age, Weight-for-age and Weight-for-height). The GDHS 2008 indicated that 28% of children under 5 years in Ghana were stunted with 9% of children less than 5 years being wasted. It also showed that 14% of children less than 5 years were underweight. The results further showed the percent distribution of variables like mothers education, mother's wealth quintile group, mothers residential status and mother's nutritional status of children less than 5 years in Ghana.

The 2008 United Nations' fourth report on world nutrition emphasized the unacceptably high prevalence of protein energy malnutrition (PEM) and anemia throughout the developing world. For children under 5 years of age, approximately 32% were stunted (height-for-age Z-score<-2 standard deviations) and approximately 9% are wasted (weight-for-height Z-score < -2 standard deviations) compared with a normal, healthy reference population. The effect of such malnutrition is exacerbated by the 3.5 billion individuals in the developing world that simultaneously suffer from iron deficiency and its resultant anemia.

In the past, the nutrition and health of school-aged children and adolescents in the developing world received little attention relative to those less than five 5 years of age. However, recent research has been more focused on school-age children because of growing evidence that 1) high prevalence and severity of PEM is sustained during these years, 2) these nutritional problems can adversely affect cognition and school/work performance. Over the past few decades, children less than 5 years in Ghana have experienced significant improvements in anthropometric measures. Over all, the proportion of children under five who are stunted decreased from 34% in 1988, to 31% in 1993, and then rose to 35% in 2003 before decreasing to 28% in 2008.The proportion of underweight children decreased from 23% in 1988 and 1993 to 14% in 2008.The proportion of children who are wasted decreased over the past five years. These statistics are obtained from anthropometric data of children under 5 years. The World Health Organization (WHO) endorses standards that classify child's health as problematic of a child's weight-, height-, or Body mass index (BMI) – for

-age are more than two standard deviations below the median in the reference population from the United States.

WHO, recognizes that the standards are appropriate only for children under 11 and are most valid for children under five years of age. Increasingly, researchers recognize that a standard based on U.S children may be inappropriate in developing world contexts, particularly given the problem of the obesity in American children.

Our understanding of the nutritional status is compromised by difficulties posed by the different standards of classifications. Typically, the 1993 and 2008 DHS used different child growth standards for under 5 years nutritional status of children.

Some of the previous studies failed to utilize certain variables in building statistical and mathematical models to identify how certain key characteristics of the mother (variables) can perform in classifying the nutritional status of young children. It is against this background that we seek to employ discriminant analysis as a technique in under 5 years anthropometric classification in Ghana.

1.6 Objectives of the Study

1.6.1 General Objective

The general objective of the study is to build an effective Discriminant Function for anthropometric nutritional classification of children (0 - 5 years) in Ghana.

1.6.2 Specific Objectives

- To classify nutritional status indicator variables using a scale of relative importance on the basis of which a parsimonious function can be developed.
- To determine the linear combinations of the variables that maximizes the difference(s) between (or within) the groups compared; stunting, wasting and underweight.
- To order variables with respect to their contribution to the classification accuracy of interest.

- To investigate the discriminant function using Logistic Regression.

1.7 Thesis Outline

The study is organized into five chapters. Chapter one consists of background of the study, problem statement, objectives of the study as well as background information of the study area.

Chapter two discusses a review of related literature of the study while chapter three deals with the methodology of the study, definitions of variables of the study, classifications of nutritional standards of children under five years based on child growth standards of WHO, derivation of the discriminant function as well as inferential procedures of discriminant analysis.

Chapter four consists of empirical results from exploratory data analysis and results from classification and logistic regression.

The summary, discussion of findings and recommendations are presented in chapter five.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Much research has been done on nutritional status of children under five years. A review of related literature studies are discussed in the following subsections.

2.2 Definition of Nutrition

Nutrition is the science that studies the interactions between living organisms and food. Human nutrition includes the study of nutrients and other substances found ingestion, digestion, absorption, transport, metabolism, interaction, storage, and excretion of nutrients by the body. In a broader sense, the study of nutrition also includes various psychological, sociological, cultural, technological, and economic factors that affect the foods and dietary patterns chosen by an individual.

Nutrition education is a critical component of most major health promotion and disease prevention programs. Research indicates that behavioral change is directly related to the amount of nutrition education received (Black 2008). Nutrition Education involves the communication of nutrition-related information that will equip individuals, families, and communities to make appropriate food choices. The media remain the primary source of nutrition information in the United States. Thus, nutrition education also focuses on discriminating between credible and non-credible sources of nutrition information. Nutrition messages and programs must be culturally relevant and specific to the target group. Registered dietitians are the professionals who are specifically trained to deliver information on food and nutrition.

2.2.1 Nutritional Assessment

A nutrition assessment is an in-depth evaluation of both objective and subjective data related to an individual's food and nutrient intake, lifestyle, and medical history. Once the data on an individual is collected and organized, the practitioner can assess and evaluate the nutritional status of that person. The assessment leads to a plan of care, or intervention, designed to help the individual either maintain the assessed status or attain a healthier status. The data for a nutritional assessment falls into four categories: anthropometric, biochemical, clinical, and dietary. We shall however look at anthropometric in detail since it forms the focus of our study.

2.3 Anthropometry

Anthropometry in physical anthropology refers to the measurement of the human individual for the purposes of understanding human physical variation. Today, anthropometry plays an important role where statistical data about the distribution of body dimensions in the population are used to optimize Changes in life styles, nutrition and ethnic composition of populations leading to changes in the distribution of body dimensions (e.g., the obesity epidemic), and require regular updating of anthropometric data collections.

2.3.1 Brief History of Anthropometry.

Alphonse Bertillon (1853), gave this name in 1883 to a system of identification depending on the unchanging character of certain measurements of parts of the human frame. He found by patient inquiry that several measures of physical features, along with dimensions of certain bones or bony structures in the body, remain fairly constant throughout adult life (*www.wikipedia.org*).

Anthropometric was first used in the 19th and early 20th century in criminalistics, for identifying criminals by facial characteristics. Francis Galton was a key contributor as well, and it was in showing the redundancy of Bertillon's measurements that he developed the statistical concept of correlation.

With more measurements of hopefully independent variables, a more precise identification could be achieved, which could then be matched against photographic evidence. Certain aspects of this philosophy would also go into Galton's development of fingerprint identification as well. An-thropometry, however, gradually fell into disfavor, and it has been generally supplemented by the superior system of finger prints. Bertillonage exhibited certain defects which were first brought to light in Bengal.

In Bengal, measurements were already abandoned by 1897, when the finger print system was adopted throughout British India. Three years later England followed suit; and as the result of a fresh inquiry ordered by the Home Office, finger prints alone were relied upon for identification.

Anthropometric studies are today conducted for numerous different purposes. Academic anthropologists investigate the evolutionary significance of differences in body proportion between populations whose ancestors lived in different environmental settings.

The US Military has conducted over 40 anthropometric surveys of U.S. Military personnel between 1945 and 1988, including the 1988 Army Anthropometric Survey (ANSUR) of men and women with its 240 measures. Statistical data from these surveys, which encompassed over 75,000 individuals.

In 2001, the UK conducted the largest sizing survey using scanners up to date. Since then there have been several national surveys which have followed in the UK's pioneering steps, notably these are Size USA, Size Mexico & Size Thailand, the latter are still ongoing. Size UK showed that the nation had got taller and heavier, but not as much as many had expected. Since 1951 when the last women's survey had taken place the average weight for women had gone up from 62 to 65 kg.

2.3.2 Design Tools for Anthropometric Data Bases

Numerous organizations around the world routinely conduct anthropometric surveys of different populations and organize the information into databases. In the surveys that look to accurately represent the compositions of these populations (henceforth referred to as "reference populations"), the subjects are sampled based on demographic variables such as their age, gender, race, ethnicity, etc. The detailed anthropometric and demographic information and the large amounts of information contained in the databases make them valuable design tools. Designers may employ either of the following two approaches when using an anthropometric database:

Performing accommodation analyses directly on the reference population and extending the results to the target user population. This approach is faulty on three counts:

a) it requires assuming that the reference and target populations are similarly composed, as measured by the distributions of demographic variables

b) it neglects the impact of temporal changes in the reference population anthropometry, and

c) it fails to consider other reasons (e.g., high fitness levels and the absence of pregnant women military populations) for possible differences in anthropometric distributions.

Utilizing various techniques (e.g., principal components analysis, the regression with residual variance methodology) to extrapolate the relationships found to exist in the reference population anthropometry to the target user population. Doing so allows for the estimation of user population anthropometry; accommodation analyses may be carried out on these estimates.

2.3.2.1 Proportionality Constant

Proportionality Constants were one of the earliest methods developed to predict human anthropometry. They are typically calculated by taking a large sample of anthropometric data and determining either the mean or 50th percentile ratio of the length of each measure of interest to stature. Drillis and Contini were among the first to publish mathematical ratios of many body dimensions to stature.

These values have been extensively used as a design tool because they provide a means of estimating the lengths of many body segments while knowing only the stature of an individual. An example design process using proportionality constants works like this:

1. Determine which body dimension can be used to most accurately predict adjustability levels. For example, trochanteric height (leg length) may be used to predict seat height adjustability range on a stationary exercise bike.

2. Determine the cutoff percentiles for the desired accommodation level for the artifact. If 95% accommodation is desired, the 2.5th and 97.5th percentiles are used.
3. Find the accommodation range. To do this, multiply the proportionality constant of the body dimension determined in step 1 by the stature corresponding to the low cutoff percentile determined in step 2. This provides the smallest body dimension length that will be accommodated. Multiply the same proportionality constant by the stature corresponding to the high cutoff percentile to obtain the largest body dimension that will be accommodated. The difference between these two lengths is the accommodation range.

Other methods of design tools includes: population model, statistical tools and Hybrid models.

2.3.3 NCHS/WHO Standards and their Implications.

From 1978 onwards, The National Center for Health Statistics (NCHS) reference has been used to assess anthropometric nutritional status in children. The new WHO standards, developed on the results of the Multicentre Growth Reference Study and published in 2006, were designed to replace them (WHO, 2006).

The introduction of these new standards has consequences on the determination of nutritional status at individual and population level. When plotting weight against height for the cut-off of weight-for-height - 3 Z-score, for NCHS reference and WHO standards, it is clear that WHO standards identify more children with a weight-for-height less than - 3 Z-score. This varies according to a child's height. Less difference is seen when weight-for-height - 2 Z-score is used. If weight-for-height less than 70% of the median is used, WHO standards tend to identify less children below this cut-off than NCHS reference .Moreover, studies based on a limited number of surveys indicated that a switch in weight-for height Z-score from the NCHS reference to the new WHO standards would have little effect on the overall prevalence of Global Acute Malnutrition (GAM) or Moderate Acute Malnutrition (MAM),but will result in a significant increase in the prevalence of Severe Acute Malnutrition(SAM).

The relationships between the NCHS and WHO-based indices did not appear to be different in populations living in different locations, and there was no consistent difference in the median age of children classified as cases by the case-definitions. During the discussion, it was noted that the WHO standards appear to classify children who were previously identified as moderately mal-nourished by NCHS reference as severely malnourished: when weight is plotted against height, the cut-off of 80% of the Median NCHS reference is close to the cut-off of -3Z-score WHO standards while the cut-off of 70% of the median NCHS/WHO, UNICEF and SCN Informal Consultation on transitioning to WHO Growth Standards: Implications for Emergency Nutrition Programmes reference is close to the cut-off M4 Z-score WHO standards. It was suggested that new cut-offs of - 4 Z-score and - 3 Z-score to define moderate and severe malnutrition, respectively, be used with the WHO standards to remain consistent with the percentage of the median NCHS reference. On the other hand, using the WHO standards with the commonly used thresholds of -3 Z-score for defining severe acute malnutrition represents a shift towards a more preventative model of treatment, catching and treating cases earlier than is currently done. It was also emphasized that when examining risk of mortality and weight for height, there is gradual exponential increase in mortality risk when weight-for-height decreases. This pattern is found in both the NCHS reference and the WHO standards. There is no clear threshold indicated where mortality risk increases.

As the new WHO standards are adopted, operational factors in the management of acute Malnutrition including financial and human resources need to be taken into consideration as well. In this study, we adopt the new WHO child growth standards.

2.4 Concept of Classification.

Classification is the grouping together of similar objects. If each object is characterized by p variables, classification can be performed according to rational criteria. Depending on the criteria used, an object could potentially belong to several classes.

2.4.1 Brief History of Classification

According to Dodge (2008), classifying the residents of a locality or a community according to their sex and other physical characteristics is an activity that date back to ancient times. The Hindus, the ancient Greeks and the Romans all developed multiple typologies for human beings. The oldest comes from Galen (129 - 199 AD).

Later on the concept of classification spread to the fields of biology and zoology. Many authors attempted to develop some methods of classifications, but the true development of classifications methods coincides with the advent of computer.

Classification methods can be divided into two large categories; classification and cluster analysis, one based on probabilities and the other one not. The first category contains, for example, discriminant analysis which is the focus of our study. The second category of classification is generally grouped under cluster analysis which is outside the scope of our study.

2.5 Discriminant Analysis

2.5.1 Brief History of Discriminant Analysis.

Dodge (2008) indicated that some of the ideas associated with discriminant analysis go back to around 1920. The English statistician Karl Pearson (1857-1936) proposed what was called the coefficient of racial likeness (CRL), a type of intergroup distance index. The CRL was studied extensively by G. M. Morant (1899-1964) in the 1920s.

In the 1920s, a study of another distance index started in India; to be formalized by P.C. Mahalanobi: (1893-1972) in the 1930s.

Fisher (1936) carried the idea of multivariable intergroup distance was translated to that of a linear composite of variables derived for the purpose of two-group classification. The distance and variable composite ideas appeared in print prior to Fisher's seminal discriminant analysis article in 1936 ("The use of multiple measurements in taxonomic problems," which appeared in Annals of Eugenics). At the suggestion of Fisher, M. M. Barnard applied two-group (predictive) discriminant analysis in a 1935 study involving seven Egyptian skull characters. Rao (1948) studied the extension of two-group classification to multiple groups. Many other extensions and refinements of Fisher's ideas have appeared since the 1940s.

Although the initial study of discriminant analysis involved applications in the biological and medical sciences, considerable interest was aroused by statisticians/methodologists in areas of study such as business, education, engineering, and psychology. The potential for the application of discriminant analysis in education and psychology (and in other areas of study?) may be attributed to methodologists associated, in one way or the other, with Harvard University during the 1950s and 1960s.

The writings about discriminant analysis for the rest of the three or four decades focused on the prediction of group membership, labeled predictive discriminant analysis (PDA) in this current study. In the non-behavioral sciences, this focus has continued to this day. In the view of some methodologists, the study of structure (through LDFs) in the context of MANOVA has considerable potential for substantive theory exploration and development. As important as such study may be considered, its use has been very limited in applied research settings over the past four decades.

2.5.2 Overview of Discriminant Analysis.

Description of group separation, in which linear functions of variables (discriminant functions) are used to describe or elucidate the differences between two or more groups is termed as Descriptive Discriminant analysis (DDA).The goals of DDA is to identify the relative contribution of the p variables to separation of the groups (Rencher, 1998).

Prediction or allocation of observation to groups, in which linear or quadratic functions of variables (classification functions) are employed to assign an individual sampling unit to one of the groups, is termed Predictive Discriminant Analysis (PDA).

In PDA, group membership is known prior to the analysis and the sole purpose of the analysis is to derive the predictive function. A predictive analysis is possible in many situations where prior designation of groups exists (e.g., product purchasers versus non-purchasers: heavy half versus light half market segments- innovators versus non-innovators; successful versus non-successful new product ideas, etc.). Again, the research objective is to predict using the set of independent variables, and not to classify consumers of unknown group membership.

Tukey (1969) studied optimization based on chance, creates a degree of fit, but in the case of the predictive analysis, this fit may be upward biased and not representative of the real world.

Seo et al. (1995) discussed the effects of non-normality on dimensionality tests in a DDA context.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter discusses the source of data, Definition of variables as well as the discriminant function analysis.

3.2 Sources of Data

The data used for this study was secondary data from the 2008 Ghana Demographic and Health Survey (GDHS) from the Head Office of the Ghana Statistical Service. The study population was children under five years. Valid measurements of height and weight of 2,992 aged five years and below were used to compute three anthropometric indices; weight-for-age, height - for -age and weight -for- height which are used to define the nutritional status of children based on a reference standard score by WHO.

3.3 Variables

The variables considered are:

- xi = Child's Body Mass Index (CBMI)
- x2 = Vaccination Status of Child (VSC)
- x3 = Mother's Highest Educational level (MHEL)
- x4 =Literacy of the Mother(LOM)

- x_5 = Child lives with whom(CLW)
- x_6 =Wealth Index of the Mother(WIM)
- x_7 =Occupation of the Mother (OCM)
- x_8 =Age of the Household Head(AGHH)
- x_9 =Sex of the Household Head(SEXHH)
- x_{10} =Antenatal Care from Government Hospital(ANCG)
- x_{11} =Antenatal Care from Private Hospital(ANCP).

The study seeks to determine whether these variables can be used to classify children (less than 5yrs) as being Stunted/Chronically malnourished or Normal, Wasted/Acutely malnourished or Normal and better still Underweight or Normal.

Empirically, the model for our study is given as;

$$y_i = a_1x_1 + a_2x_2 + \ldots + a_px_p \tag{3.1}$$

Where x_1, x_2 $\cdots x_p$ are vectors of independent variables consisting of measured attributes of both the child and the mother. y_i is the grouping variable. $i = 1,\ 2\ 3$.

Thus,

$$y_1 = \begin{cases} 0, \text{ Stunted} \\ 1, \text{ Normal} \end{cases}, \quad y_2 = \begin{cases} 0, \text{ Wasted} \\ 1, \text{ Normal} \end{cases} \text{ and } \quad y_3 = \begin{cases} 0, \text{ Underweight} \\ 1, \text{ Normal} \end{cases}$$

Our study seeks to determine how these measured attributes of the children and their mothers can be used to distinguish among each of the groups (y_1, y_2 and y_3)

3.4 Fisher's Discriminant Function Analysis

Discriminant analysis is a powerful descriptive and classificatory technique that is used to describe characteristics that are specific to distinct groups (called descriptive discriminant analysis); and Classify cases (i.e. individuals subjects, participants) into pre-existing groups based on similarities between that case and the often cases belonging to the groups (sometimes called predictive discriminant analysis).

3.4.1 Data Requirements for Discriminant Analysis

Tinsley and Brown (2000) outlined a number of data requirement for discriminant analysis. We shall outline these requirements in the light of our study.

Discriminant Analysis requires that a data set contains two or more mutually exclusive groups.

Scores on two or more variables for each case in the group may be constructed on the basis of demographic characteristics.

Selected variables should satisfy requirements for ordinal level measurements.

Non-dichotomous nominal variables can be used but they must be dummy coded into dichotomous categories Examining our study, the three distinct populations of children (less than 5 years) thus, Stunted, Wasted, and Underweight have satisfied the grouping requirement. Again, it is clear from our study that demographic characteristics such as age, sex, occupation etc. have contributed to the construction of the three groups. The three groups (Stunted, Wasted and Underweight) are categorical, based on a continuous scale measurement. It has been recoded to dichotomous groups. Once these data requirements have been satisfied, it forms a very good basis for which we can perform discriminant analysis. To ensure reliability, stability and generalizability of the results, we have decided to adopt the total number of children (2,992) in the data set for our study as small samples usually influence Statistical results. (Huberty, 1975).

It is recommended that sample size should be at least ten times the number of discriminator variables. Stevens (1996) also argued that the ratio of cases to variables should be more on the order of 20 to 1.

Despite these requirements our study adopted the total population of children under five years in Ghana

3.4.2 Assumptions

Statistical Considerations in discriminant analysis have to do with distributional assumptions concerning observations, measures of separation among groups, algorithms for carrying out both stages of the discriminant analysis and the study of the properties of proposed algorithms. Vi-are observed as;

a). Independence of Observations;

b). Multivariate normality

c). Homogeneity of covariance matrices

These assumptions have been in existence and used as major requirement and the basis for the application of most statistical methods. Contemporary research has shown that violations of some of these assumptions have little influence on effect size (s).

Current evidence suggests that discriminant analysis is robust with respect to violation of assumptions of multivariate normality and of homogeneity of covariance matrices (Stevens, 1996). Discriminant analysis is especially robust to violations of the assumption of homogeneity of the covariance if the ratio of the largest group, n, divided by the smallest group, n, is less than 1.5 (Stevens, 1996). We can also test for violation of homogeneity of the covariance matrices assumption by using the Box's M statistic in SAS and SPSS. When violated, Kleczka (1980) noted that the worst consequence is that cases are more likely to be classified into the group with the greater dispersion.

3.4.2.1 Examination of the Underlying Assumptions

We shall explore the data in chapter four to examine the assumptions underlying discriminant analysis. For normality, we shall use the boxplot, the Kolmogorov Smirnov test as well as skewness and kurtosis. For homogeneity of covariance matrices, we shall use the Box's M statistic. Tabacehnick and Fidel (1996) showed that the assumption of independence is examined by consulting published correlation matrices of the discriminant variables. Generally, correlations having an absolute value of less than 0.3 have no interpretative value and do not violate the assumption.

3.5 Missing Values

Missing values distorts statistics produced by the analysis and leads to results that do not generalize. We can handle the issue of missing values by either ignoring them or replacing all missing values with the mean. The later has disadvantage of reducing the within-group heterogeneity on that variable, however, we shall still replace all missing values with the mean because ignoring missing values can be very serious especially when variables on which information is missing is related to other variables included in the data set (Tinsley and Brown, 2000).

3.6 Error Rate Estimation

We shall examine the performance of our classification function by consideration one of the methods below.

First, the re-substitution method; this method yields what Hills (1966) calls the apparent error rate. Letting P_1 and P_2 denote the misclassification probabilities of erroneously assigning an observation to group i (G_1) when the observations comes from group j (G_2), the $\hat{p}_1$ and $\hat{p}_2$ are simply the sample proportions of misclassified observations. The estimates are consistent but can be severely optimistically biased.

Secondly, the hold out method procedure splits the total sample into two. One subsample is used to construct classification rule, and the other is used for validation.

Another method is the U-method or Cross-Validation which was first used by Iachenbruch and Mickey (1968). It holds one observation at a time, estimates the discriminant function based on $N_1 + N_2 - 1$ observation and classifies the hold out Observation. This process is repeated until all observations are classified. This methods yields almost unbiased estimates of the misclassification probabilities.

It is also called the leave-one-out and inaccurately referred to as a Jackknife procedure. The fourth method is the boot strap method. This seems to combine the best features of cross-validation and the resubsituation method. It has small variance and almost unbiased.

Generally, the error rates determined by using the methods above are;

1. The Optimum error rate; the rate, which would hold if all parameters were known.
2. The Actual error; the rate that holds for a classification rule under consideration when it is used to classify all possible future samples.
3. The Apparent error rate; the rate we obtain by resubstituting the training samples and determining the misclassifications.

In this study, the cross-validation (leave-one-out) method is used because it yields almost unbiased estimates of the misclassification probabilities.

3.7 Derivation of the Discriminant Function

Suppose $\mathbf{X}' = [X_1, X_2, . . .Xp]$ is a p-dimensional random vector with mean vector $\mu = [\mu_1, \mu_2, \ldots\ldots \mu p]$ and covariance matrix Σ.

The discriminant function analysis attempts to find a discriminating function based on measurements obtained from some correlated variables of X. This function will be a linear combination of the X and a p-dimensional vector of weights; So that given a sample from a population

$$Y = a_1x_1 + a_2x_2 + \ldots\ldots a_px_p \tag{3.2}$$

Let

$$\mathbf{X} = [\, X_1\, X_{2i} \ldots\, Xpi] \tag{3.3}$$

The two group and three group functions were expressed in Adebanji (2000), as shown below in section [3.8] and [3.9] respectively.

3.8 Classical Two-Group Linear Function

Let

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \tag{3.4}$$

Be observations from normally distributed populations. The objective is to obtain the vector A_p so that Y_j will be the discriminant score of the J^{th} observed individual

Let

$$Y = a_1x_1 + a_2x_2 + \ldots + a_px_p = a^{'}x \tag{3.5}$$

Given two vectors X_1 and X_2 with sizes n_1 and n_2 respectively, where $n = n_1 + n_2$, $\bar{Y_1}$ and $\bar{Y_2}$ the means of the groups with standard deviation of 1.

$\mathbf{D} = \bar{Y_1} - \bar{Y_2}$

$\Rightarrow \mathbf{D} = a^{'}\bar{x_1} - a^{'}\bar{x_2} = a^{'}[\bar{x_1} - \bar{x_2}]$

$$\therefore \mathbf{D} = a^{'}d \tag{3.6}$$

$$\mathbf{D}^2 = [\bar{Y_1} - \bar{Y_2}]^2 \tag{3.7}$$

$$\mathbf{W} = S(Y_1) + S(Y_2) \tag{3.8}$$

$\Rightarrow \mathbf{W} = S(a^{'}x_1 + S(a^{'}x_2) = a^{'}[S_1^* + S_2^*]a$

$$\therefore \mathbf{W} = a^{'}Sa \tag{3.9}$$

Equation (3.9) is the within groups sum of squares.

$S_1^* = (n_1 - 1)Var(x_1)$ and $S_2^* = (n_2 - 1)Var(x_2)$

It is assumed that S_1^* and S_2^* are equal and both are estimates of a common population variance matrix.

$$Var(x_1) = (n_1 - 1)^{-1}\sum(x_{1i} - \bar{x_1})(x_{1i} - \bar{x_1})^{'} \tag{3.10}$$

$$Var(x_2) = (n_2 - 1)^{-1}\sum(x_{2i} - \bar{x_2})(x_{2i} - \bar{x_2})^{'} \tag{3.11}$$

Next is to divide W by $\sum(n_g - 1) = n_1 + n_2 - 2$ to obtain the variance covariance matrix.. the best linear discriminant is that which maximizes the ratio $\frac{D^2}{W}$ with respect to a. This can be obtained from the Lagrangian expression.

$$L(a) = D^2 - \theta(W - C) \tag{3.12}$$

It can be shown from calculus that differentiating $L(a)$ with respect to a and equating the partial derivative to zero gives $D = Sa$.

Hence

$$\mathbf{a} = S^{-1}d \tag{3.13}$$

This gives p normal equations which can be solved for p unknowns provided S is nonsingular. The discriminant function is

$$Y = a_1x_1 + a_2x_2 + \ldots + a_px_p \tag{3.14}$$

$$\bar{Y}_1 = a_1\bar{x}_{11} + a_2\bar{x}_{21} + \ldots + a_p\bar{x}_{p1} \tag{3.15}$$

$$\bar{Y}_2 = a_1\bar{x}_{12} + a_2\bar{x}_{22} + \ldots + a_p\bar{x}_{p2} \tag{3.16}$$

If it is assumed that the cost of misclassification is the same for each group, that is; $C(1|2) = C(2|1)$ and that $\bar{Y}_1 > \bar{Y}_2$, the familiar halfway classification rule stipulates that an individual with observed vector x be classified into group 1 if

$a^{'}x > \frac{1}{2}(\bar{Y}_1 + \bar{Y}_2)$

$d^{'}S^{-1}x > \frac{1}{2}[a^{'}\bar{x}_1 + a^{'}\bar{x}_2]$

$d^{'}S^{-1}x > \frac{1}{2}[d^{'}S^{-1}\bar{x}_1 + d^{'}S^{-1}\bar{x}_2]$

$$d^{'}S^{-1} > \frac{1}{2}d^{'}S^{-1}[\bar{x}_1 + \bar{x}_2] \tag{3.17}$$

and into group 2 if

$$d^{'}S^{-1} \leq \frac{1}{2}d^{'}S^{-1}[\bar{x}_1 + \bar{x}_2] \tag{3.18}$$

The discriminant function can also be used to assign an individual with observation into one of K groups, where $K > 2$. The procedure is to compute the linear discriminant scores,

$$W_{ij} = \left[x^{'} - \frac{1}{2}(\bar{x}_1 + \bar{x}_2)^{'}S^{-1}(\bar{x}_1 - \bar{x}_2)\right] \tag{3.19}$$

And assign x to group 1 if $W_{ij} > 0$ for $i \neq j$. since $W_{ij} = -W_{ij}$, only k-1 linearly independent scores can be computed.

The cut-off point "c" can be chosen in various ways. Sometimes it is chosen so that the number misclassified of some of the K training samples is as small as possible. If p variables used in the discrimination are normally distributed, and if their covariance matrices are the same in say, two groups/populations, then a frequently used cut-off point is

$$C = ln\left[\frac{\hat{\pi}_2}{\hat{\pi}_1}\right] \tag{3.20}$$

Hence $\hat{p}_i(g)$ is some estimate of $\pi(g)$ the a prior probability that an individual to be classified come from $\pi(g)$. If two populations or groups have normal distribution with equal covariance matrices, then equation (3.20) is the best possible classification rule in the sense that expected probability of misclassification is as small as possible, that is

$$P = \hat{\pi}_1 p(2|1) + \hat{\pi}_2 p(1|2) \tag{3.21}$$

is minimized. Where $p(2|1)$ is the probability of misclassifying an individual from π_2 when little is known about the relative population sizes, it is usual to set

$$\hat{\pi}_1 = \hat{\pi}_2 = \frac{1}{2}, \text{ so that } ln\left[\frac{\hat{\pi}_2}{\hat{\pi}_1}\right] = C = 0 \tag{3.22}$$

3.9 Classification into Three Groups

We extend the two groups case to the three groups case as our study involves three population or groups.

Let $k = 3$, then we need 2 classification scores.

Compute

$$W_{12} = \left[x' - \tfrac{1}{2}(\bar{x_1} + \bar{x_2})'\right] S^{-1}(\bar{x_1} - \bar{x_2})$$
$$W_{13} = \left[x' - \tfrac{1}{2}(\bar{x_1} + \bar{x_3})'\right] S^{-1}(\bar{x_1} - \bar{x_3})$$

The decision rule is:

If $W_{12} > 0$, $W_{13} > 0$, then assign x to group 1

If $W_{12} < 0$, $W_{13} > W_{12}$, then assign x to group 2

If $W_{12} < 0$, $W_{12} > W_{13}$, then assign x to group 3

Classification is such that $(x_1 < x_2 < x_3)$

3.10 Inferential Procedures in Discriminant Function Analysis

Inferential procedures for tests of hypotheses, significance of discrimination, homogeneity of groups and significance of canonical correlation was carried out in this study. The major test are as follows:

3.10.1 Hotelling's T^2 Test

Under the normality assumption and equality of covariance matrices, the hypothesis that the P-variables have no discriminating power (i.e. the derived function has no discriminating power) can be stated as,

H_0 =The function has no discriminating power.

This can be tested with the statistic

$$T^2 = \left[\frac{n_1 n_2}{n_1 + n_2}\right] D^2 \tag{3.23}$$

Where n_1 and n_2 are the group sample sizes and D^2 is defined earlier in equation (3.7). Equation (3.23) has an approximate F distribution (Adebanji 2000) with

$$F = \left(\frac{n_1 n_2 D^2}{(n_1 + n_2)P}\right) \Big/ \left(\frac{D}{n_1 + n_2 - p - 1}\right) \tag{3.24}$$

3.10.2 Wilk's (1932) Lambda Test

It is the ratio of the determinant of within group variation to total variation. If it has values close to 1, this indicates that almost all of the variability in the discriminator variables is due to within-group difference, and values close to 0 indicates that almost all of the variability in the discriminator variable is due to group difference (Brown and Wicker, 2000) . When it is rejected, inspection of the univariate -ratios may suggest which of the elements of the vector variables are contributing most to the discrimination of groups (Adebanji,2000).

The test $H_0 : \mu_i = \mu \; i = 1,\ 2, \ldots\ g$ is a test of homogeneity of groups, or even as a test of the treatments represented by the groups; it was formulated by Wilks (1932) in terms of the distribution of a ratio of determinants. The test statistic is usually denoted as lambda(λ) and defined as;

$$\lambda = \frac{|\mathbf{W}|}{|\mathbf{T}|} \tag{3.25}$$

Where $\mathbf{W}$ =within groups sum of squares and $\mathbf{T}$ =Total sum of squares

3.10.3 Box's M Test

This is a test criterion for the null hypothesis of the equality of group dispersion matrices, extended from the work of Bartlett's (1947), and has been presented by Box (1949). Many research workers prefer to ignore the test for homogeneity group dispersion on the grounds that Wilks' test is probably fairly robust under departures from its assumptions (Adebanji, 2000).

Box defines the test criterion for M

$H_0 = \triangle_k = \triangle,\ k = 1,\ 2,\ 3 \ldots g$

$$\mathbf{M} = (N-g)log_e|\triangle_w| - \sum_{k=1}^{g}(N_K - 1)log_e|\triangle_k| \tag{3.26}$$

Where $\triangle_k$ is the dispersion estimate for the K^{th} sample and $\triangle_w$ is the pooled group estimate based on W.

3.10.4 The Chi Squared Test

Bartlett (1947) derived a procedure for testing the significance of canonical correlations. He defines the test statistic lambda.

$$\lambda = \prod_{i=1}^{p^2}(1-\lambda_i) \tag{3.27}$$

The null hypothesis that our two discriminant functions are Unrelated in the classification of the nutritional status of the children (less than 5years) would be tested by the function of equation (3.27) and that is approximately distributed as a Chi-Squared with p_1p_2 degrees of freedom .That is

$$\chi^2 = -[n - 0.5(p_1 + p_2 + 1)]\, log_e\lambda \tag{3.28}$$

As a rule of thumb, authority frequently treats canonical correlations of 0.3 or less as trivial (Adebanji, 2000).

3.10.5 Likelihood Ratio Test

This is a test of significance for each effect in a fitted logistic regression model. These tests compare the likelihood function of a full model to that of the model in which only the predicted effect has been dropped. Small p - values indicate that the model has been improved significantly by the corresponding effect. This test is analogous to the forward stepwise method in discriminant function analysis which selects the variable that maximizes the Wilks's lambda.

3.11 Logistic Regression (LR)

Logistic regression is a powerful statistical tool that can be used to model several responses that arises in various fields of study.It can be used to model binary responses (eg yes or no), ordinal responses (eg normal, mild and severe) and also nominal. Other literature on logistic regression can be gotten from Cox and Sneil (1989).

In this study, our dependent variable is binary. Thus

$$y_1 = \begin{cases} 0, \text{ Stunted} \\ 1, \text{ Normal} \end{cases}, \quad y_2 = \begin{cases} 0, \text{ Wasted} \\ 1, \text{ Normal} \end{cases} \text{ and } \quad y_3 = \begin{cases} 0, \text{ Underweight} \\ 1, \text{ Normal} \end{cases}$$

It means that the response of an individual child can take one of the possible values above basd on the measurement(variables) of that child.

Suppose further that x is a vector of explanatory variables and $\pi = p_r(Y = 1/x)$ is the response probability to be modeled. the linear logistic model is of the form

$$\log it(\pi) \equiv= \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta' x \tag{3.29}$$

Where α is the intercept and β is the vector of parameters

This is quite similar to discriminant analysis except that Logistic Regression does not make use of the underlying assumptions of normality and homogeneity of covariance matrix.

Gilbert(1968) showed that Fisher's linear discriminant function is optimal when two populations have multivariate normal distributions with equal covariance matrices, and has been shown to be relatively robust to departures from normality. But in severe cases of non-normality - for example when the variables are binary - it might be worth considering an alternative approach. One commonly employed involves the use of a logistic function to model directly the probability of an observation being a member of each group. As stated earlier in our objectives, we shall cross examine our results from the 2008 DHS data on discriminant function analysis and logistic regression and see how inferentially, the behaviour of the variables.

The parameters may be estimated by maximum likelihood. Day and Kerridge (1967) showed that this method of discriminant analysis has optimal properties under a wide range of assumptions about the underlying distributions, indicating those relevant when both continuous and categorical variables are used to describe each member of the training set.

CHAPTER 4

DATA PRESENTATION AND RESULTS

4.1 Introduction

This chapter discusses some preliminary analysis of the study. It also presents the empirical results of the research. The chapter further outlines the analysis under the following headings; Canonical Correlations Coefficients of the DFA, Relative importance of the variables, the Wilks's Lambda and Chi Square Test, DFA Coefficients for both standardized and unstandardized. The Box M'S Test, Group Centroids, Fisher's Linear Discriminant function and Evaluation/efficiency of the classification function. The chapter further shows results of cross examination between DFA and LR.

4.2 Assumptions

The graphs in Appendix A indicated that the data are not normally distributed for the three populations thus, Stunting, wasted and underweight. Nonetheless, we indicated that we adopt the Fisher's Linear Discriminant function which is robust to non normality (see assumptions in chapter three). Our Box M Tests is significant for each of the three groups indicating unequal covariance matrices but this is not a problem (Stevens, 1996).

4.3 Preliminary Analysis

A cross-tabulation of selected variables was carried out on the data: Sex of household head by nutritional status of the children, wealth index of the mother by the nutritional status of the children and the distribution of nutritional status of the children by regions in Ghana.

4.3.1 Sex of Household Head (Sex of HH) by Nutritional Status of Children Under Five

In this analysis, computations were among sex in each group for each category (Stunted, Wasted and Underweight).

We can see from table 4.1 that among the normal group under stunted for height, male headed

Table 4.1: Nutritional Status of Children under Five by Sex of Household Head

Nutritional Status of Child		Sex of Household Head				Total
		Male		Female		
		Coun	Percent	Count	Percent	
Stunted for height	Normal	407	72.9	151	27.1	558
	Stunted	1376	75.6	445	24.4	1821
Wasted for height	Normal	379	72.6	143	27.4	522
	Wasted	1404	75.6	453	24.4	1857
Underweight	Normal	707	74.9	237	25.1	944
	Underweight	1076	75	359	25	1435

Source: Computed from the 2008 GDHS Data Set.

households appear to have more children (72.9%) who are normal than female headed household (27.1%). For the malnourished group under stunted for height, male headed households recorded the highest (75.6%) of malnourished children than the female headed household (24.4%).

But it is clear that there are more male headed households (1783) than female headed households (596).

For the normal group under wasted, there are more male headed households (1783) than female headed households (596) and this has reflected in the Relative percent recorded for both sexes. Generally, the males headed households have done very well than the female headed households. This could be attributed to the fact that many households in Ghana are headed by males. Household head is the bread winner of the house.

4.3.2 Wealth Index by Nutritional Status of Children Under 5 Years

Wealth Index		Stunted for		Wasted for		Underweight	
		Normal	Stunted	Normal	Wasted	Normal	Underweight
Poorest	Count	153	626	114	665	301	478
	Percent	27.4	34.4	21.8	35.8	31.9	33.3
Poorer	Count	113	421	106	428	212	322
	Percent	20.3	23.1	20.3	23	22.5	22.4
Middle	Count	89	293	79	303	141	241
	Percent	15.9	16.1	15.1	16.3	14.9	16.8
Richer	Count	100	300	111	289	163	237
	Percent	17.9	16.5	21.3	15.6	17.3	16.5
Richest	Count	103	181	112	172	127	157
	Percent	18.6	9.9	21.5	9.3	13.5	10.9
Total		558	1821	522	1857	944	1435

Source: Computed from 2008 GDHS Data Set

The GDHS 2008 used the SPSS factor analysis procedure to compute the wealth index. This procedure first standardized the indicator variables (calculating the Z scores); then the factor coefficient scores (factor loadings) are calculated; and finally, for each household, the indicator values are multiplied by the loadings and summed up to produce the household index value (Rustein and Johnson 2004). For analytical purposes, quintiles were used which are based on distribution of household population rather than on distribution of households.

The cut-off point in the wealth index at which to form the quintiles are calculated by obtaining a weighted frequency distribution of households, the weights being the products of the number the de jure members of the households and sampling wage of the household, thus the distribution represents the national household population where each member is given the wealth index score for his/her household. The persons are then ordered by the score thus poorest, poor, middle, richer, and richest. These indices were already determined and coded into the GDHS 2008 dataset.

Table 4.2 shows that for the stunted for height category, the poorest group has recorded more stunted (malnourished) children (34.4%) and the richest group recorded the least (9.9%). Similarly, the wasted category also recorded 35.8% wasted (malnourished) children under the poorest group and the richest group recorded the least wasted (malnourished) children (9.3%).

The underweight category follows the same trend. That is, poorest recorded high underweight children (33.3%) and the richest group recorded the least (10.9%).Generally the poorest group in the country tend to have their children exposed to long term effects of malnutrition which reflects in the cumulative growth of their children (Stunted) and low body mass in relation to body height or length wasted (thin).

43.3 Distribution of Nutritional Status of Children Under Five by Regions in Ghana.

Table 4.3: Distribution of Nutritional Status of Children Under Five by Regions in Ghana.

Nutritional Status			WR	CR	GA	VR	ER	AS	BA	NR	UE	UW
Stunted For Height	Normal	Count	45	33	67	50	41	98	51	89	35	49
		Percent	8.1	5.9	12	9	7.3	17.6	9.1	15.9	6.3	8.8
	Stunted	Count	168	130	146	162	155	273	188	283	138	178
		Percent	9.2	7.1	8	8.9	8.5	15	10.3	15.5	7.6	9.8
Wasted for Height	Normal	Count	56	42	68	47	50	94	185	52	24	35
		Percent	10.7	8	13	9	9.6	18	10	10	4.6	6.7
	Wasted	Count	157	121	145	165	146	277	91	320	149	192
		Percent	8.5	6.5	7.8	8.9	7.9	14.9	9.6	17.2	8	10.3
Under-weight	Normal	Count	100	71	85	91	106	147	148	130	59	64
		Percent	10.6	7.5	9	9.6	11.2	15.6	10.3	13.8	6.2	6.8
	Under-weight	Count	113	92	128	121	90	224	-	242	114	163
		Percent	7.9	6.4	8.9	8.4	6.3	15.6	-	16.9	7.9	11.4

Source: Computed from 2008 GDHS Data Set. (Note: Computation of percentages is among regions.) - = Missing

The distribution of nutritional status of the children (less than 5 years) by regional basis showed that among the stunted (malnourished) group under the stunted for height category, Northern Region (NR) recorded the highest malnourished cases (15.5%), Central Region (CR) recorded the least (7.1%). We can also see from Table 4.3 that for the wasted (Thin) category, Northern Region again recorded the highest (17.2%) percentage of malnourished (Thin) children and Central Region (CR) recorded the least (6.5%).

For underweight category, the situation is the same, 16.9% of underweight children are from the Northern Region whilst 6.3% which is the least come from the Eastern Region (ER).

Generally, the trend seem to be that the Northern Region tends to have more malnourished children (less than 5 years) and Central Region (CR) seem to have the least number of malnourished children (less than 5 years). This could possibly be due to the fact that the poverty level in the Northern Region is higher and also polygamy has contributed to higher number of children in the north.

4.4 Results from Discriminant Function Analysis (DFA)

4.4.1 Group Centroids

The absolute magnitude of the group (Nutritional status of the child) Centroid indicates the degree to which a group is differentiated on a function, and the signs of the centroid indicate the direction of the differentiation.

Table 4.4 shows that the function discriminates Normal from stunted (chronically malnourished).

Table 4.4: Group Centroids of Discriminant Function

Nutritional Status		Function 1
	Normal	0.12
Stunted for	Stunted	-0.43
	Normal	0.06
Wasted for	Wasted	-0.46
	Normal	0.05
Underweight	Underweight	-0.42

Normal scored at the positive end of the function and stunted (chronically malnourished) score at the negative end.

The function clearly discriminates Normal from wasted (acutely malnourished) with normal scoring positive and wasted (acutely malnourished) scoring the negative end of the function.

For the underweight, which reflects on the body composition in relation to weight of the child, the function differentiates normal from the underweight in a similar fashion as in the two cases above.

The group centroid represents the mean discriminant score of members of a group on a given discriminant function.

For classification and prediction purpose, the discriminant score of each case (e.g an individual child) is compared to each group centroid and the probability of group membership is calculated. The closer a score is to a group centroid, the greater the probability that the case belong to that group.

4.4.2 Relative Importance of the Discriminatory Variables

We order the discriminating variables in terms of their relative contribution to the discriminant function. Only selected variables from stepwise discriminant analysis are ordered for each of the three categories (Stunted for height, wasted for height and underweight).

Table 4.5: Canonical Correlations

Nutritional Status	Discriminators	Function
Stunted for Height	CBMI	0.85
	VSC	-0.58
Wasted for Height	CBMI	1.00
Underweight	CBMI	1.00

CBMI-Child Body Mass Index, VSC- Vaccination Status of child

An inspection of Table 4.5 revealed that Body Mass Index of the child contributes highly (0.85) to a child being normal or stunted (chronically malnourished), followed by the Vaccination status of the child (-0.58). This is an indication that the BMI of the child differentiates a child as being Normal or Chronically malnourished.

Interestingly, Body Mass Index of the child appears to have a perfect loading for wasted for height and underweight. This shows that the Body Mass Index is important in predicting the status of the child (either wasted or underweight). The Canonical Correlations provides an indication of the practical value of our discriminant function.

4.4.3 The Discriminant Model.

When classifying in K groups, k-1 discriminant functions are required. This study requires classifying the nutritional status of each category into two groups.

For the stunted for height category, we require two groups, thus stunted or normal, thus one discriminant function is required. See definition of the dependent variables in chapter three,

Thus the model for stunted for height (y_1) is;

$$y_1 = 0.82x1 - 0.53x_2 \tag{4.1}$$

x_1 = CBMI and x_2 = VSC

Wasted for height y_2 is

$$y_2 = x_1 \tag{4.2}$$

Where x_1 = CBMI

Underweight (y3) is

$$y_3 = x_1 \tag{4.3}$$

It must however be noted that the standardized Coefficients were used because these are used to determine the comparative relations of discriminator variables to the functions hence the above models are used for descriptive discriminant analysis. Table 4.6 below contains results for both standardized and unstandardized discriminant coefficients.

Table 4.6: Discriminant Function Coefficients

Status	Function	Standardized Coefficient			Unstandardized Coefficient		
		x	a_2	constant	a_1	a_2	constant
y_1	1	0.82	-0.53	-	0	-0.81	0.38
y_2	1	1	-	-	0	-	-0.38
y_3	1	1	-	-	0	-	-0.37

a_1 and a_2 are the discriminators

For prediction purposes, we could build our discriminant function with the unstandardized coefficients and consider our derived Fisher's linear discriminant function.

It is observed that in determining whether a child (less than 5 years) is normal or stunted/chronically malnourished, knowledge of the BMI and the vaccination status of the child are important in classifying the child. However, vaccination status of the child has an inverse relationship with the discriminant function.

The BMI of the child is the only variable that discriminates clearly the status of the child being wasted/acutely malnourished or normal .The same applies to underweight or normal.

For the prediction purposes, we could use the unstandardized Coefficients for the discriminant function.

$$y_1 = -0.81x_1 + 0.39 \qquad (4.4)$$

$$y_2 = -0.38 \qquad (4.5)$$

$$y_3 = -0.37 \qquad (4.6)$$

4.4.4 Fisher's Linear Discriminant Function for the Three Categories

Table 4.7: Fisher's Linear Discriminant Function for the Three Categories

	Stunted For Height		Wasted For Height		Underweight	
Variable	Normal	Stunted	Normal	Wasted	Normal	Underweight
CBMI	0	3.42E-05	0	-2.53E-05	0	-1.36E-05
VSC	1.98	2.41	-	-	-	-
Constant	-1.18	-2.79	-2.13	-0.30	-0.21	-2.20

Table 4.7 contains the Fisher's linear discriminant function which is robust to normality assumption. We can perform our classification for each of the three categories of child nutritional status.

4.4.5 Classification Table (Confusion Matrix) Analysis

In this study, the confusion matrix has traditionally ended the discriminant analysis table. This table was obtained from the analysis of our 2008 DHS data. This was subjected to a variety of analysis that may be directed towards answering the following questions.

1. What level of overall classification is expected from chance?

2. Which groups are best classified by the discriminant function?

This study examined the classification accuracy of the discriminant function for the three categories of child nutritional status.

Critically, we shall cross-validate the results of the discriminant function by using the leave-one-out method since one of our objectives is to build a function for future classifications.

Cross -validation procedures are necessary to determine how much shrinkage in the hit rates (percent of correct predictions) can be expected when classifying cases that were not used to derive the function (Tinsley and Brown,2006).

4.4.5.1 Evaluation of the Performance of the Classification Function for Stunted Category

Table 4.8: Classification (Confusion Matrix) Results for Stunted Category

Data Type		Nutritional Status	Predicted Group Membership		Total
			normal	stunted	
Original Data	Count	normal	2326	5	2331
		stunted	657	4	661
Original Data	%	normal	99.8	0.2	100
		stunted	99.4	0.6	100
Cross validation	Count	normal	2326	5	2331
		stunted	657	4	661
Cross validation	%	normal	99.8	0.2	100
		stunted	99.4	0.6	100

The classification table above reveals that the percentage of stunted for height correctly classified is 77.9% out of the 2992 cases using the derived discriminant function. This observed classification was found to be significant at 0.005 level (x^2 = 17.24, *df* = 2). Within the stunted category normal was best classified (99.8%) indicating more normal children are classified into the normal group in the stunted category. We performed leave-one-out procedure or cross validation to determine the shrinkage and found that 77.9% of the cross validated classifications were correct. This is almost the same result we got from the developmental sample.

Thus, the use of the discriminant function derived from the developmental sample to classify nutritional status of independent samples of GDHS data on children under five can be expected to result in approximately 80% of the cases being correctly classified. This indicates a very good performance of the discriminant function for the stunted category.

Table 4.9: Classification (Confusion Matrix) Results for Wasted Category

Data Type		Nutritional Status	Predicted Group		Total
			normal	wasted	
Original Data	Count	normal	2766	0	2766
		wasted	226	0	226
Original Data	%	normal	100	0	100
			100	0	100
Cross-validation	Count	normal	2766	0	2766
		wasted	226	0	226
Cross-validation	%	normal	100	0	100
		wasted	100	0	100

The results of wasted for height showed 100% correct classifications for normal children (less than 5 years) predicted as normal, and 100% misclassification of wasted/acutely malnourished as normal. No misclassification for normal. The overall hit rate was 92.4% indicating that 92.4% out of 2992 cases were correctly classified by the derived discriminant function. This observed classification was significant at 0.05 level(x^2 = 9.37, df = 1) see table 4.9. The cross validated results also classified 92.4% of the grouped case correctly. There was 100% correct classification

Table 4.10: Classification (Confusion Matrix) Results for Underweight Category.

Data Type		Nutritional Status	Predicted Group		Total
			normal	underweight	
Original Data	Count	normal	2646	0	2646
		underweight	346	0	346
Original Data	%	normal	100	0	
		underweight	100	0	100
Cross-validation	Count	normal	2646	0	
		underweight	346	0	346
Cross-validation	%	normal	100	0	100
		underweight	100	0	100

for normal children predicted as normal and 100% misclassification for underweight children predicted as normal. However, there was no misclassification for normal as underweight and correct classification for underweight as normal.

Generally, the function correctly classified 88.4% of the grouped cases. Cross-validated results also classified 88.4% of the grouped cases correctly. Again, the BMI of the child as a discriminator variable has performed well as variable that can be used to discriminate the child's status as normal or underweight.

4.5 Comparison Between Classification Results of Discriminant Function Analysis (DFA) and Logistic Regression (LR)

Even though, we have cross validated our results in the discriminant function to assess the performance of the function. We have chosen to build a logistic regression model on the same set of variables we used for classifying the nutritional status of the children in discriminant analysis with the forward likelihood ratio method (Forward stepwise).

The logistic regression model examines the weights (discriminators in DFA) to determine their effects on the outcome variable. Chi square test was used to assess the significance of the weights.

Logistic regression enabled us to examine the odds of falling into an outcome category given a one-unit change in a specific predictor. These odds are useful when interpreting which independent variables (discriminators in DFA) provide relevant information in predicting group membership in the outcome variable. Higher values of odds ratios (exp (B)) indicate associated independent variables have greater odds of falling into the baseline category (that is the one coded 1 verse a 0 in a dichotomous outcome).

We adopt the definition of the dichotomous dependent variable in chapter three for our three nutritional categories (stunted, wasted and underweight), given the same variables in section 3.3.

The forward stepwise method was used to run the selection procedure for all the three nutritional categories.

4.5.1 Variables in the Equation for Stunted Category

An inspection of the variables selected revealed that only BMI of child (hw73) and VSC (h10) were significant at 5% alpha level. Further inspection of the classification table indicates that 99.6% of normal cases were correctly classified. The overall performance of the logistic model in classifying children as stunted or normal was 78.5%.These results appeared to be consistent with the discriminant function performance for stunted for height category. The step summary of the logistics regression model showed that 78.8% of BMI cases were correctly classified. See table 4.11 below.

Table 4.11: Classification Table for Stunted Category (LR)

Observed		Predicted		
		stunted for height		Percentage Correct
		normal	stunted	
stunted for height	normal	281	1	99.6
	stunted	76	0	0
Total %				78.5

This revealed that logistics regression is more effective in terms of individual performance of the independent variables. Clearly, the Chi Square test was significant at 5% level of significance, showing that CBMI and VSC are significant in the model for both classification and prediction purposes.

The odds ratio for the BMI of the child indicates that children whose BMI are low, are 1 times more likely to be classified as stunted/chronically malnourished. This result is also consistent with the discriminant function for the stunted for height which has 0.82 as loading for CBMI as it contributes highly to the discrimination of the outcome category.

4.5.2 Variables in the Equation for Wasted Category

Similarly, the logistic procedure for wasted for height category captures only the BMI of the child (hw73) as the only independent variable. The classification table shown below indicates 98.4% correct classification for normal and 87.5% were correctly classified as wasted. It is observed that the overall classification accuracy of the logistic regression model is 97.2%. It must however be noted that the logistic regression model has done well in differentiating between normal and wasted than the discriminant function for the wasted category which was 92.4%. Further inspection of

Table 4.12: Classification Table for Wasted Category (LR)

Observed		Predicted		
		wasted for height		Percentage Correct
		normal	wasted	
wasted for height	normal	313	5	
	wasted	5	35	87.5
Overall Percentage				97.2

the logistic regression output revealed that the weight of the model is -0.08 with an odds ratio of 0.927 (see Table 4.16).

Comparatively, the discriminant loading for the CBMI in the discriminant function is 1.00 (See table 4.14).This clearly shows that whiles the weight in the logistic regression has an inverse effect on whether a child is normal or wasted. The discriminant loading indicates positive perfect effect on the outcome variable. The odds ratio indicates that children with low BMI are more than half times more likely to fall in the wasted/acutely malnourished group.

4.5.3 Variables in the Equation for Underweight

The BMI of the child was significant (at 5% level of alpha) in differentiating the child being normal or underweight. Comparatively, the discriminant function derived earlier also captured the BMI only. 98.7% of normal children were correctly classified with only 10% of underweight children correctly classified. Generally, the logistic model was 88.8% correct in classifying a child as normal or underweight comparatively our discriminant function correctly classified 88.4% of grouped cases. Further inspection of the analysis indicated that the weight of the logistic regression for this category is -0.01 (see Table 4.16), correspondingly, our discriminant loading is 1.00 (See table 4.14).

Indicating that, this is the only contributing discriminatory variable in determining whether a child is normal or underweight. The odds ratio of 0.99 (see Table 4.16) shows a higher probability of a child being classified as underweight.

Table 4.13: Classification Table for Underweight Category (LR)

Observed		Predicted		
		underweight		Percentage Correct
		normal	underweight	
underweight	normal	314	4	98.7
	underweight	36	4	10
Total %				88.8

4.6 Summary of Canonical DFA and Weights in LR

Table 4.14: Summary of Canonical DFA and Weights in LR

	Standardized Canonical DF			Logistic Regression		
Variables	stunted	Wasted	underweight	stunted	wasted	underweight
CBMI	0.82	1	1	0	-0.08	-0.01
VSC	-0.53	-	-	-	-	-
CONSTANT	-	-	-	-0.03	-14.03	-2.41
ODDS	-	-	-	1	0.93	0.99

Table 4.15: Evaluation of the Performance of Classification Function of Discriminant Function and Logistic Regression

Variables	% of correct classification (stunted)	% of correct classification (wasted)	% of correct classification (underweight)
DFA	77.9	92.4	88.4
LR	78.5	97.2	88.8

Clearly, the logistic regression as an investigative tool has proven that our classification functions were effective. However, the performance of the logistic regression classification model has proved a bit more effective than the DFA classification which is the norm, since we are using logistic regression as an investigative tool. Even though, the CBMI and VSC are significant in

Table 4.16: Confidence Interval (C.I) for Odds Ratio for Logistic Regression

Nutritional Status	95% C.I		Odds Ratio	Significant Variables	
	Lower	Upper		CBMI	VSC
Stunted For	0.999	1	1	0	-
Wasted For	0.895	0.961	0.927	-0.075	-
Underweight	0.991	0.996	0.993	-0.007	-

both cases of DFA and LR, our summary table above shows that the odds ratio of the model for stunted category is enough to determine the classification of the outcome category for stunted for height indicating that CBMI and VSC are significant, but they are not necessary in determining the outcome category.

LR as an investigative tool showed the confidence interval for all the categories. If repeated samples are taken, we are 95% confident that between 0.991 and 0.996 of the children will be classified as stunted/chronically malnourished.

Similarly, wasted for height has between 0.895 and 9.61 as C.I and underweight has between 0.991 and 0.996.

4.7 Evaluation of Selected Variables

To investigate the dominance of the CBMI in the discriminant model, we standardized the selected variables as listed in section 3.3 using the SPSS. These standardized variables were used to run the discriminant model and we had the BMI still dominating. Even though WHO recognized the BMI as a key determining factor in nutritional classification, in this study we used the discriminant analysis as an applied tool which justified WHO's assertion.

CHAPTER 5

SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Introduction

This chapter summarizes empirical results from the analysis as well as draws conclusions and make recommendations that are expected to contribute to knowledge and impact on existing policies in Ghana.

The concept of anthropometric indicators of children under five years is very broad. This study is just one of the several methods of using selected attributes of both children (less than 5 years) and their mothers to classify the anthropometric indicators of children (less than 5 years) in Ghana using the 2008 GDHS Data set.

The preliminary analyses revealed that the incidence of malnutrition in children under five in Ghana is more pronounced in the Northern Region for all the three nutritional deficiency categories, which is, stunting, wasted and underweight.

The study identified the poorest group in Ghana as having more malnourished children under five than those who belong to the richest group.

5.2 Summary

The performance of the LR classification model has proved a bit more effective than the DFA classification (see table 4.15).

The odds ratio of the LR model indicated that, CBMI and VSC are significant, but they are not necessary in determining the stuntedness of a child under five in Ghana.

LR as an investigative tool showed the confidence interval for all the categories. If repeated samples are taken, we will be 95% confident that the proportion of children who will be classified as stunted/chronically malnourished will be between 0.991 and 0.996.

Similarly, wasted for height has a proportion between 0.895 and 9.61 as C.I and underweight has a proportion between 0.991 and 0.996.

5.3 Conclusions

The discriminant function clearly identified the Body Mass Index of the child as the major determining factor in classifying the nutritional status of children (less than 5 years) in Ghana. The Body Mass Index (BMI) measures the body mass in relation to body height or length and classifies current nutritional status. Stuntedness is an indication of linear growth retardation and cumulative growth deficits. For stuntedness, BMI contributes greatly.

Vaccination status of the child (VSC) impacts negatively on the child being stunted or normal. Children who are wasted for height are considered thin and represents failure to receive adequate nutrition. The BMI of the child clearly discriminates a child as being wasted/acutely malnourished or normal. The probability of a child being classified as stunted/chronically malnourished is certain (1) if the child has low score of BMI. Again, children (less than 5 years) are 0.927 times more likely to be classified as wasted which reflects in the acutely malnourished condition of the child with a low score of BMI, the probability of a child being classified as underweight is 0.993. Underweight combines both acute and chronic malnutrition condition of the child.

5.4 Recommendations

The nutritional status of the children should further be broken down into three groups for each category. That is:

$$\text{Stunted for height}, (y_1) = \begin{cases} 2 & \text{Stunted} \\ 3 & \text{Severely stunted} \\ 1 & \text{Normal} \end{cases}, \text{Wasted for height}, (y_2) = \begin{cases} 2 & \text{Wasted} \\ 3 & \text{Severely wasted} \\ 1 & \text{Normal} \end{cases}$$

$$\text{and Underweight}, (y_3) = \begin{cases} 2 & \text{Underweight} \\ 3 & \text{Severely underweight} \\ 1 & \text{Normal} \end{cases}$$

Sufficient controls should be exercised in the discriminant function for the three categories. For instance, the BMI can be controlled to see the behavior of the other discriminator variables.

To further investigate the discrimination between the known categories (stunted, wasted and underweight) with similar techniques such as classification trees and Neural Networks.

BIBLIOGRAPHY

Adebanji, A. O. (2000). Discriminant Analysis as a technique for classifying Nigerian Households into Poverty Levels and Income Classes; Unpublished Thesis.

Bartlett, M, S. (1947). Multivariate Analysis, *Journal of royal Statistical Society* B, 9, pp.176-197

Bernard M. M. (1935). The secular variation of skull characters in four series of Egyptian skulls. *Annals of Eugenics,* 6, 352-371.

Black, R. E., L. H. Allen, et al. (2008). Maternal and child under nutrition: global and regional exposures and health consequences. The Lancet.

Cox and Sneil (1989). Analysis of binary data (2nd ed.). Chapman and Hall, London.

Day and Kerridge (1967). A general maximum-likelihood discriminant, *Biometrics*, 23,313-323 de Onis, M. and M. Blossner (2003). The World Health Organization Global Database on Child Growth and Malnutrition: methodology and applications. *International Journal of Epidemiology*. deOnis, M., C. Monteiro, et al. (1993).The worldwide magnitude of protein?energy malnutrition: an overview from the WHO Global Database on Child Growth. Bulletin of the World Health Organization.

Devereux, S. S. Vaitla, et al. (2008). Seasons of Hunger; Fighting cycles of quiet Starvation among the world's rural poor, A Hunger Watch Publication.

Dillon, W. R. and Goldstein M. (1984). Multivariate Analysis, Methods and Applications. pp 372 - 379. John Wiley and Sons. New York.

Dodge .Y. (2008).The Concise Encyclopedia of Statistics. Springer, Germany.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics,* pp. 179-188.

Ghana Statistical Service (2008). Ghana Demographic and Health Survey Final Report 2008.

Ghana Statistical Service (2008).Ghana Demographic and Health Survey Data set 2008.

Ghana Statistical Service (GSS) and Macro International Inc. (MI) 1999.Ghana Demographic and Health Survey 1998. Calverton, Maryland: GSS and MI.

Ghana Statistical Service (GSS), Noguchi Memorial Institute for Medical Research (NMIMR), and ORC Macro. 2004. Ghana Demographic and Health Survey 2003. Calverton, Maryland: GSS, NMIMR and ORC Macro.

Gilbert E. S.(1968); Discrimination using qualitative variables. *Journal of the American Statistical Society* 63, 1399-1412.

Hills, M. (1966); Allocation rules and their error rates.*Journal of the Royal Statistical Society*, B28, p. l.

History of Anthropometry, 25th April 2010 (www.wikipedia.org).

Hotelling Harold (1935). The most predictable criterion, *Journal of educational Psychology*, 26, 139-142.

Howe and Deverenx (2004). Famine intensity and magnitude scales; "A proposal for an instrumental definition of famine" Disaster 28(4): 353-372.

Huberty C. J. (1975). The stability of three indices of relative variable contribution in discriminant analysis. *Journal of Experimental Education* pp. 59-64.

Kleczka, W. R. (1980). Discriminant Analysis. Beverly Hills, CA. Sage.

Lachenbruch and Mickey (1968); Estimation of error rates in discriminant analysis, *Technometrics*, 10, 1-11.

Mahalanobis, P. C. (1930). Test and measures of group divergence.*Journal and Proceedings of the Asiatic Society of Bengal*, 26, 541-588.

Pellefier et al. (1993).Epidemiologic evidence for a potentiating effect of malnutrition on child morality. *American Journal of Public Health*. 83(8).1130-1133.

Rencher, A. C. (1998). Multivariate Statistical Inference and Applications. New York, Wiley.

Roa, C. R. (1948). The utilization of multiple measurement in problems of biological classification, *Journal of the Royal Rustein and Johnson (2004). The DHS wealth index: DHS comparative report number 6. Calverton, Maryland Statistical Societ*, B. 10, 159-193.

Seo, T., Kanda, T., and Fujikoshi, Y. (1995). The effects of nonnormality on tests for dimensionality in canonical correlation and MANOVA models. *Journal of Multivariate Analysis*. pp. 325-337.

Stevens, J. (1996). Applied multivariate Statistics for the social Sciences, 3rd ed. Mahwah, N. J.; Lawrence Erlbaun.

Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics (3rd ed.). New York: Harper Collins.

The World Bank. 2003. Ghana Poverty Reduction Strategy 2003-2005: An agenda for growth and prosperity. Vol. 1: Analysis and policy statement. Washington D.C.: The World Bank.

The United Nations (2008). IRIN News: *Health care article.* www.irinnews.org/report. August 5 2008. United Nations.

Tinsley and Brown (2000).Handbook of Applied Multivariate Statistics and Mathematical Modeling. Academic Press, USA.

Tukey, J. W. (1969). Analyzing data: Sanctification of detection work. *American Psychologist*, 24, Pp83-91.

Waterlow, J. C. (1972). Classification and definition of protein calorie malnutrition. *British Medical Journal* 3:566-569.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance, *Biometrika,* 24, 471-494.

World Health Organisation (WHO) Multicenter Growth Reference Study Group. 2006. WHO child growth standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Geneva. World Health Organisation.

APPENDIX A

Exploratory Data Analysis

```
EXAMINE VARIABLES=hw71
  /PLOT BOXPLOT STEMLEAF HISTOGRAM NPPLOT
  /COMPARE GROUP
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
EXAMINE VARIABLES=hw72
  /PLOT BOXPLOT STEMLEAF HISTOGRAM NPPLOT
  /COMPARE GROUP
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

Explore

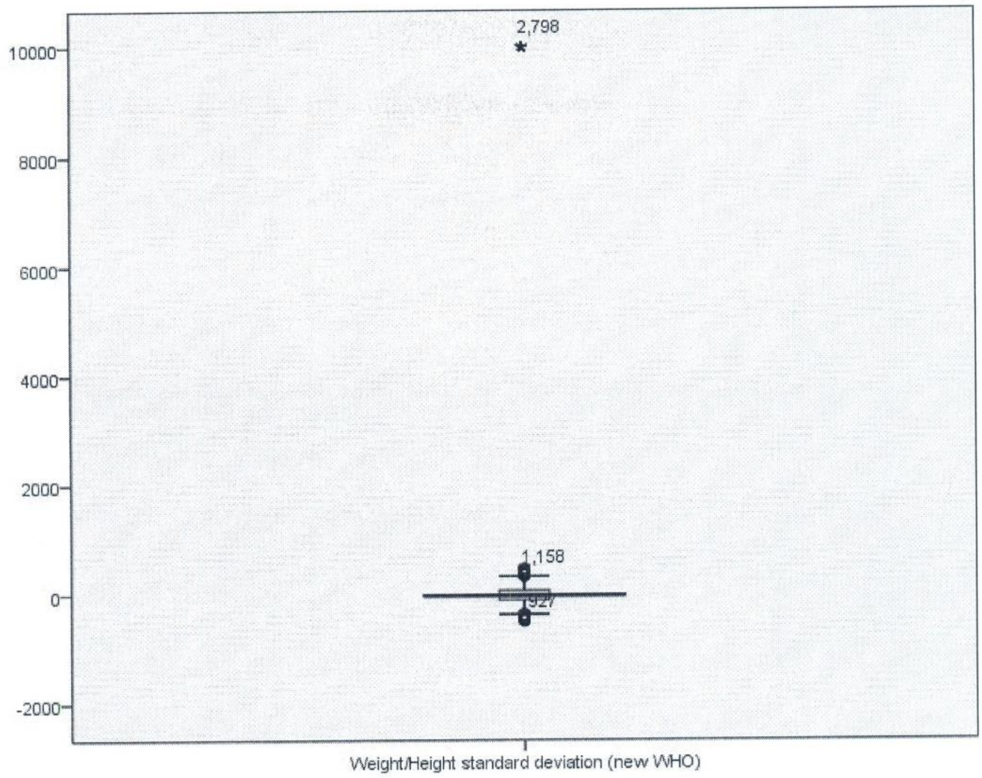
Descriptives

			Statistic	Std. Error
Weight/Height standard deviation (new WHO)	Mean		597.31	48.318
	95% Confidence Interval for Mean	Lower Bound	502.57	
		Upper Bound	692.06	
	5% Trimmed Mean		126.27	
	Median		-21	
	Variance		5.93E+06	
	Std. Deviation		2.43E+03	
	Minimum		-493	
	Maximum		9998	
	Range		10491	
	Interquartile Range		175	
	Skewness		3.594	0.049
	Kurtosis		10.972	0.097

Tests of Normality

	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Weight/Height standard deviation (new WHO)	0.473	2538	0	0.306	2538	0

a. Lilliefors Significance Correction



2,798
10000
8000
6000
4000
2000
0
-2000
1,158
Weight/Height standard deviation (new WHO)

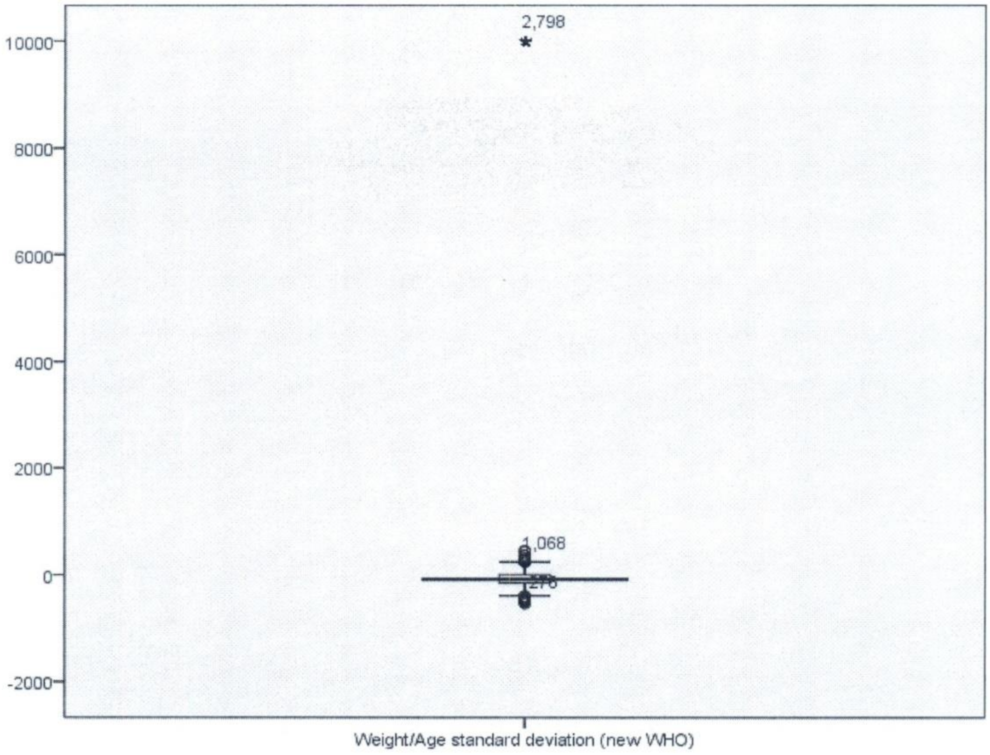
Descriptives

			Statistic	Std. Error
Weight/Age standard Deviation (new WHO)	Mean		548.13	48.553
	95% confidence Interval for Mean	Lower Bound	452.92	
		Upper Bound	643.34	
	5% Trimmed Mean		72.03	
	Median		-77	
	Variance		5.98E+06	
	Std. Deviation		2.45E+03	
	Minimum		-529	
	Maximum		9998	
	Range		10527	
	Interquartile Range		157	
	Skewness		3.598	0.049
	Kurtosis		10.992	0.097

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Weight/Age standard deviation (new WHO)	0.481	2538	0	0.3	2538	0

a. Lilliefors Significance Correction




```
EXAMINE VARIABLES=hw70
  /PLOT BOXPLOT STEMLEAF HISTOGRAM NPPLOT
  /COMPARE GROUP
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

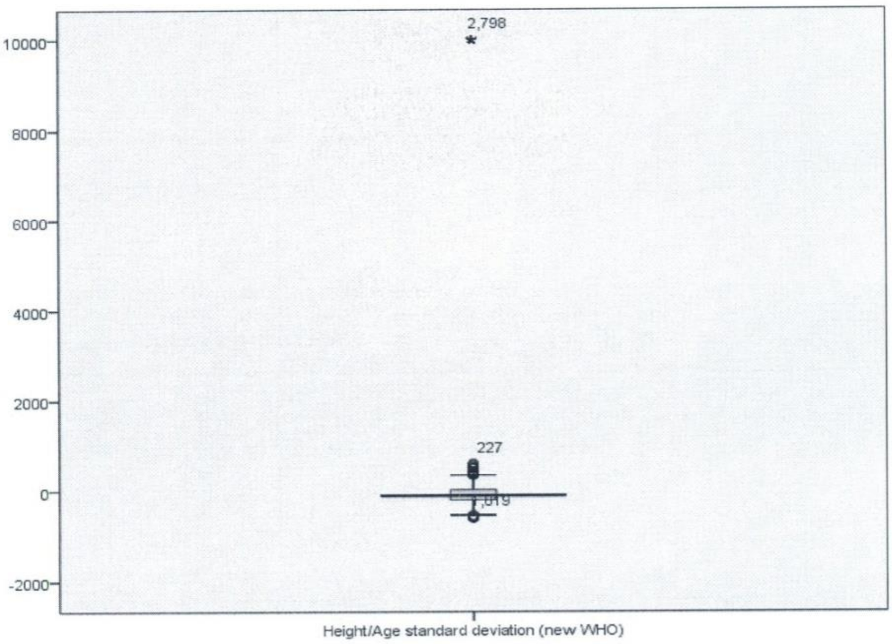
Descriptives

			Statistic	Std. Error
Height/Age Standard deviation (new WHO)	Mean		525.08	48.719
	95% Confidence Interval for	Lower Bound	429.55	
		Upper Bound	620.62	
	5% Trimmed Mean		52.2	
	Median		-107	
	Variance		6.02E+06	
	Std. Deviation		2.45E+03	
	Minimum		-590	
	Maximum		9998	
	Range		10588	
	Interquartile Range		222	
	Skewness		3.586	0.049
	Kurtosis		10.938	0.097

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Height/Age standard deviation (new WHO)	0.461	2538	0	0.315	2538	0

a. Lilliefors Significance Correction


2,798
10000
8000
6000
4000
2000
0
-2000
227
Height/Age standard deviation (new WHO)

APPENDIX B

Output of Discriminant Analysis

```
DISCRIMINANT
 /GROUPS=stunted(0 1)
 /VARIABLES=hw73 h10 v151 v152 v155 b9 v106v716 v190m57g m57r
 /ANALYSIS ALL
 /METHOD=WILKS
 /FIN=3.84
 /FOUT=2.71
 /PRIORS SIZE
 /HISTORY
 /STATISTICS=MEAN STDDEV BOXM COEFF RAW TABLE CROSSVALID
 /PLOT=COMBINED SEPARATE MAP
 /CLASSIFY=NONMISSING POOLED MEANSUB.
```

	Function
	1
BMI standard deviation (new WHO)	0.849
Ever had vaccination	-0.582
Wealth index[a]	-0.093
Antenatal care: government health post/CHPS[a]	0.092
Age of household head[a]	-0.09
Sex of household head[a]	-0.061
Highest educational level[a]	-0.05
Literacy[a]	-0.033
Respondent's occupation[a]	0.014

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions.

Variables ordered by absolute size of correlation within function.

a. This variable not used in the analysis.

Stepwise Statistics

Step	Entered	Wilks' Lambda							
		Statistic	dfl	df2	df3	Exact F			
						Statistic	dfl	df2	Sig.
1	BMI standard deviation (new WHO)	0.965	1	1	356	12.847	1	356	0
2	Ever had vaccination	0.952	2	1	356	8.887	2	355	0

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

a. Maximum number of steps is 22.

b. Minimum partial F to enter is 3.84.

c. Maximum partial F to remove is 2.71.

d. F level, tolerance, or VIN insufficient for further computation.

Variables in Analysis

Step		Toleranc	F to Remove	Wilks' Lambda
1	BMI standard deviation (new WHO)	1	12.847	
2	BMI standard deviation (new WHO)	0.996	11.567	0.983
	Ever had vaccination	0.996	4.791	0.965

```
DISCRIMINANT
  /GROUPS=wasted(0 1)
  /VARIABLES=hw73 h10 v151 v152 v155 b9 v106 v716 v190 m57g m57r
  /ANALYSIS ALL
  /METHOD=WILKS
  /FIN=3.84
  /FOUT=2.71
  /PRIORS SIZE
  /HISTORY
  /STATISTICS=MEAN STDDEV BOXM COEFF RAW TABLE CROSSVALID
  /PLOT=COMBINED SEPARATE MAP
  /CLASSIFY=NONMISSING POOLED MEANSUB.
```

Structure Matrix

	Function
	1
BMI standard deviation (new WHO)	1
Ever had vaccination[a]	-0.091
Wealth index[a]	-0.073
Sex of household head[a]	-0.058
Respondent's occupational'	0.057
Antenatal care: government health post/CHPS[a]	0.055
Age of household head[a]	-0.054
Highest educational level[a]	-0.033
Literacy[a]	-0.003

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions.

Variables ordered by absolute size of correlation within function.

a. This variable not used in the analysis.

Classification Statistics

Classification Processing Summary

Processed		2992
Excluded	Missing or out-of-range group codes	0
	At least one missing discriminating variable	0
Used in Output		2992

Wasted for height			Predicted Group Membership		Total
			normal	wasted/malnourished	
Original	Count	normal	2766	0	2766
		wasted/malnourished	226	0	226
	%	normal	100	0	100
		wasted/malnourished	100	0	100
Cross-validateda	Count	normal	2766	0	2766
		wasted/malnourished	226	0	226
	%	normal	100	0	100
		wasted/malnourished	100	0	100

Classification Results[a,b]

a. Cross validation is done only for those cases in the analysis. In cross validation each case is classified by the functions derived from all cases other than that case.

b. 92.4% of original grouped cases correctly classified.

c. 92.4% of cross-validated grouped cases correctly classified.

Stepwise Statistics

Variables Entered/Removed[a,b,c,d]

Step	Entered	Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	BMI standard deviation (new WHO)	0.974	1	1	356	9.503	1	356	0.002

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

a. Maximum number of steps is 22.

b. Minimum partial F to enter is 3.84.

c. Maximum partial F to remove is 2.71.

d. F level, tolerance, or VIN insufficient for further computation. variables in the Analysis

Step		Tolerance	F to Remove
1	BMI standard deviation (new Who)	1	9.503

```
 DISCRIMINAT
/GROUPS = underwght (0 1)
```

```
/VARIABLES=hw73 h10 v190 v152 v151 v106 v716 b9 m57 m57r
/ANALYSIS ALL
/METHOD=WILKS
/FIN=3.84
/FOUT=2.71
/PRIORS SIZE
/HISTORY
/STATISTICS=MEAN STDDEV BOXM COEFF RAW TABLE CROSSVALID
/PLOT=COMBINED MAP
/CLASSIFY=NONMISSING POOLED MEANSUB.
```

Discriminant

Structure Matrix

	Function
	1
BMI standard deviation (new WHO)	1
Ever had vaccination[a]	-0.073
Wealth index[a]	-0.069
Antenatal care: government health post/CHPS[a]	0.061
Sex of household head[a]	-0.05
Age of household head[a]	-0.049
Respondent's occupation[a]	0.045
Highest educational level[a]	-0.031

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

a. This variable not used in the analysis.

Classification Results

underweight			Predicted Group Membership		Total
			normal	underweight	
Original	Count	normal	2646	0	2646
		underweight	346	0	346
	%	normal	100	0	100
		underweight	100	0	100
Cross-validated'	Count	normal	2646	0	2646
		underweight	346	0	346
	%	normal	100	0	100
		underweight	100	0	100

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 88.4% of original grouped cases correctly classified.

c. 88.4% of cross-validated grouped cases correctly classified. Variables Entered/Removed[a,b,c,d]

Step	Entered	Wilks' Lambda							
		Statistic	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	BMI standard deviation (new WHO)	0.978	1	1	357	7.916	1	357	0.005

At each step, the variable that minimizes the overall Wilks' Lambda is entered.

a. Maximum number of steps is 20.

b. Minimum partial F to enter is 3.84.

c. Maximum partial F to remove is 2.71.

d. F level, tolerance, or VIN insufficient for further computation.

Stepwise Statistics

Variables in the Analysis

Step	Toleranc	F to Remove
1 BMI standard deviation (new WHO)	1	7.916

APPENDIX C

Output of Logistic Regression

```
LOGISTIC REGRESSION VARIABLES wasted
/METHOD=FSTEP(LR) hw73 h10 m57r m57g b9 v716 v190 v155 v152
v151 v106
/CONTRAST (b9)=Indicator
/CONTRAST (v190)=Indicator
/CONTRAST (v155)=Indicator
/CONTRAST (v716)=Indicator
/CONTRAST (m57r)=Indicator
/CONTRAST (h10)=Indicator
/CONTRAST (m57g)=Indicator
/CONTRAST (v151)=Indicator
/CONTRAST (v106)=Indicator
/CLASSPLOT
/PRINT=GOODFIT CORR SUMMARY CI(95)
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).

LOGISTIC REGRESSION VARIABLES underweight
/METHOD=FSTEP(LR) hw73 h10 m57r m57g b9 v716 v190 v155 v152
v151 v106
/CONTRAST (b9) =Indicator
/CONTRAST (v190)=Indicator
/CONTRAST (v155)=Indicator
       /CONTRAST (v716)=Indicator
/CONTRAST (m57r)=Indicator
/CONTRAST (h10)=Indicator
/CONTRAST (m57g)=Indicator
/CONTRAST (v151)=Indicator
/CONTRAST (v106)=Indicator
/CLASSPLOT
/PRINT=GOODFIT CORR SUMMARY CI(95)
/CRITERIA=PIN(0.05) POUT(0.10) ITERATE(20) CUT(0.5).
```


Logistic Regression

Dependent Variable Encoding

Original Value	Internal Value
normal	0
underweight	1

Block 1:Method=Forward Stepwise(Likelihood Ratio

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 2	Step	42.99	1	0
	Block	42.99	1	0

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
2	207.696a	.113	.225

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

Classification Table[a]

Observed			Predicted		
			underweight		Percentage Correct
			normal	underweight	
Step 2	underweight	normal	314	4	98.7
		underweight	36	4	10
	Overall Percentage				88.8

a. The cut value is .5

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I. for EXP(B)
								Lower
Step 2[a]	hw73	-0.007	0.001	26.081	1	0	0.993	0.991
	Constant	-2.414	0.233	107.18	1	0	0.089	

a. Variable(s) entered on step 1: hw73.

Correlation Matrix

		Constant	hw73
Step 2	Constant	1	0.632
	hw73	0.632	1

Step Summary[a,b]

Step	Improvement			Model			Correct Class %	Variable
	Chi-square	df	Sig.	Chi-square	df	Sig.		
1	42.99	1	0	42.99	1	0	88.8%	IN: hw73

a. No more variables can be deleted from or added to the current model.

b. End block: 1

Classification Table[a,b]

Observed			Predicted		
			underweight		Percentage Correct
			normal	underweigh	
Step 0	underweight	normal	318	0	100
		underweight	40	0	0
	Overall Percentage				88.8

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-2.073	0.168	152.713	1	0	0.126

Logistic Regression

Dependent Variable Encoding

Original Value	Internal Value
normal	0
wasted/malnourished	1

Block 1: Method=Forward Stepwise (Likelihood Ration

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 2	Step	214.169	1	0
	Block	214.169	1	0

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
2	36.518a	.450	.894

a. Estimation terminated at iteration number 14 because parameter estimates changed by less than .001.

Classification Table[a]

Observed			Predicted: wasted for height: normal	Predicted: wasted for height: wasted/malnourished	Percentage Correct
Step 2	wasted for height	normal	313	5	98.4
		wasted/malnourished	5	35	87.5
	Overall Percentage				97.2

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 2[a]	hw73	-0.075	0.018	17.676	1	0	0.927	0.895	0.961
	Constant	-14.033	3.285	18.248	1	0	0		

a. Variable(s) entered on step 1: hw73.

Correlation Matrix

		Constant	hw73
Step 2	Constant	1	0.992
	hw73	0.992	1

Step Summary[a,b]

Improvement			Model			Correct Class %	Variable
Chi-square	df	Sig.	Chi-square	df	Sig.		
214.169	1	0	214.169	1	0	97.2%	IN: hw73

a. No more variables can be deleted from or added to the current model.

b. End block: 1

Block 0: Begining Block

Classification Table[a,b]

Observed			Predicted: wasted for height: normal	Predicted: wasted for height: wasted/malnourished	Percentage Correct
Step 0	wasted for height	normal	318	0	100
		wasted/malnourished	40	0	0
	Overall Percentage				88.8

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-2.073	0.168	152.713	1	0	0.126