STATISTICAL ANALYSIS OF DETERMINANTS OF HOUSEHOLD FOOD INSECURITY IN POST-CONFLICT SOUTHERN SUDAN

Laila Barnaba Lokosang

Submitted in partial fulfilment of the requirements for the degree of

Master of Agriculture (Food Security),

African Centre for Food Security,

School of Agricultural Sciences and Agribusiness,

Faculty of Science and Agriculture,

University of KwaZulu-Natal,

Pietermaritzburg

DECLARATION

I, Laila Lokosang, declare that:

(i) The research reported in this work, except where otherwise indicated, is my original research.

(ii) This work has not been submitted for any degree or examination at any other university.

(iii) The data and findings of this study remain the property of the Government of Southern Sudan, which will duly receive a full digital copy of the report, data and analysis along with this work.

(iv) This work does not contain another person's data, pictures, graphs or other information, unless specifically acknowledged as being sourced from such persons.

(v) This work does not contain another author's writing, unless specifically acknowledged as being sourced from other authors. Where other sources have been quoted, then:

- their words have been re-written, but the general information attributed to them has been referenced
- where their exact words have been used their writing has been placed inside quotation marks and referenced

(vi) This mini-dissertation does not contain text, graphics or tables that are copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and the reference section.

………………………….. **……………………………**

Laila Barnaba Lokosang Date

As Research Supervisor, I agree to submission of this work for examination.

………………………….. ……………………………

Professor SL Hendriks Date

As Research Co-Supervisor, I agree to submission of this work for examination.

………………………….. **……………………………**

Dr. Shaun Ramroop Date

DEDICATION

To my beloved wife and sons.

ACKNOWLEDGEMENT

Foremost, I thank my Lord Jesus Christ for superfluous love. To Him be the glory.

I wholeheartedly thank my supervisor, Professor S. L. Hendriks, for encouragement and guidance, especially the unique and enriching research methodology course done in a relaxed workshop environment. I also present a bouquet of roses to my Co-Supervisor, Dr. Shaun Ramroop of the School of Statistics and Actuarial Science, for all the open heartedness and guidance with the statistical methodology used. It was a fulfilling experience of statistics made easy.

I owe an emotional word of thanks to the FAO/SIFSIA Programme Support Unit team in Juba and Rome for facilitating my sponsorship. From the team I recognize Dr. Elijah Mukhala (Achimwene) for continued motivation. I also thank all staff and colleagues at the African Centre for Food Security (ACFS) for according me the entire warm and friendly environment that really boosted my commitment to this work.

Finally, thanks and love to my family for bearing with me when I had to leave them for months to pursue postgraduate studies. May God reward their patience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACCRONYMS

ACFS	:	African Centre for Food Security
AIDS	:	Acquired Immune Deficiency Syndrome
ANOVA	:	Analysis of Variance
BMI	:	Body Mass Index
CAADP	:	Comprehensive African Agriculture Development Programme
CFSVA	:	Comprehensive Food Security Vulnerability Analysis
CPA	:	Comprehensive Peace Agreement
DDS	:	Dietary Diversity Score
DFID	:	Department for International Development
DHS	:	Demographic and Health Survey
FAFS	:	Framework for African Food Security
FAM	:	Food Aid Management
FANTA	:	Food and Nutrition Technical Assistance
FAO	:	Food and Agriculture Organisation of the United Nations
FAST	:	Food Access Survey Tool
FCG	:	Food Consumption Group
FCS	:	Food Consumption Scores
FEWS NET	:	Famine Early Warning Systems Network
FIVIMS	:	Food Insecurity and Vulnerability Information and Mapping Systems
FSAU	:	Food Security Analysis Unit (FAO – Somalia)
GDP	:	Gross Domestic Product
GOSS	:	Government of Southern Sudan
HDDS	:	Household Dietary Diversity Score
HDI	:	Human Development Index
HDR	:	Human Development Report
HFIAS	:	Household Food Insecurity Access Scale
HIV	:	Human Immuno-deficiency Virus
HRW	:	Human Rights Watch
IDP	:	Internally Displaced People/Population

IFAD	:	International Fund for Agricultural Development
IFPRI	:	International Food Policy and Research Institute (IFPRI)
IGAD	:	Intergovernmental Authority on Development
MDG	:	Millennium Development Goal
MICS	:	Multiple Indicator Cluster Survey
MLR	:	Multiple Linear Regression
MOH	:	Ministry of Health
NEPAD	:	New Partnership for Africa's Development
NY	:	New York
OLR	:	Ordinary Linear Regression
POM	:	Proportional Odds Model
ReSAKSS	:	Regional Strategic Analysis and Knowledge Support System
SCF-UK	:	Save the Children Fund – United Kingdom
SHHS	:	Sudan Household Health Survey
SIFSIA	:	Sudan Information for Food Security in Action
SMART	:	Measuring Mortality, Nutritional Status, and Food Security in Crisis Situations (SMART Methodology)
SPSS	:	Statistical Package for the Social Sciences
TB	:	Tuberculosis
UK	:	United Kingdom
UKZN	:	University of KwaZulu-Natal
UN	:	United Nations
UNDP	:	United Nations Development Programme
UNICEF	:	United Nations Children's Fund
USAID	:	United States Agency for International Development
VAM	:	Vulnerability Analysis (Assessment) and Mapping
WFP	:	World Food Programme
WHO	:	World Health Organisation

MAP OF SOUTHERN SUDAN

Source: Retraced by Author (not drawn to scale)

ABSTRACT

Hunger and food insecurity has remained an endemic problem in Southern Sudan for the last three decades. Lack of a “gold standard” measure for determining causes of household food insecurity is well documented in the Food Security literature and the chase is still on for universally agreed standards. However, the Comprehensive African Agriculture Development Programme (CAADP)[1] Framework for African Food Security (FAFS) has outlined four categorical measures for structured monitoring of household food insecurity, which are yet to be rolled out for implementation by country-level Food Security programmes.

Analysis of a national household survey dataset has not been done using robust logistic regression techniques for statistically determining the factors influencing food insecurity in Southern Sudan. If such attempts are made, national food security programmes and the government statistical agency are not formally made aware of the results or do not own them. Hence, the agency has continued to lack institutional capacity to adapt the tools and techniques.

This project attempts to explore the use of robust statistical techniques featuring the Ordinal Logistic Regression procedures of SPSS for analysing the Sudan Household Health Survey (2006) dataset and determine the strengths and magnitude of relationships of nineteen independent variables in predicting categories of food consumption scores. Food Security experts and international organisations, have regarded Food Consumption Scores as a proxy measure of Food Insecurity. Twelve factors were found to statistically determine food consumption. It is, therefore, ascertained that if this form of analysis were carried out immediately after the survey was completed it would have enabled prediction of the outcome of food insecurity in Southern Sudan for at least the following year. Nevertheless, the study found out that the same statistical modelling procedures could be adopted in similar national surveys. Indeed the study provides a basis for creating an institutional memory for statistical agencies to carry out similar analysis and thereby reducing data processing time without due reliance on outsourced international expertise.

---

[1] A programme of the New Partnership for Africa’s Development (NEPAD)

CHAPTER ONE

THE PROBLEM AND ITS SETTING

1.0 Introduction

The problem of food insecurity and hunger has been persistent in many parts of the world, especially among vulnerable populations. The developing countries have experienced the bulk of the problems, which are manifested in extreme forms of hunger and chronic and acute malnutrition. Rosegrant and Cline (2003) predict that "Global food security will remain a worldwide concern for the next 50 years and beyond." They attribute this trend to (a) decreasing crop yields in many parts of the world due to scarcity of water and drop in research and infrastructure investment; (b) climatic changes and; (c) HIV and AIDS. (Webb and Rogers, 2003) caution that "there is evidence that the momentum for change initiated in the 1990s has stalled and progress will likely be harder to achieve in the future". These alarm and others by specialists in the field, point to the fact that food insecurity is a reality that must be tackled and committed by all that it takes, just like the menace of HIV and TB pandemics. Aroused by this daunting reality, heads of states and governments gathered in Rome Italy, in an historic World Food Summit – convened in1996 – issued the Rome Declaration on World Food Security (FAO, 1996). An extension of the Rome Declaration was the formulation of the Millennium Development Goal number 1[2] (UN, 2000), which places hunger and poverty reduction at the top of seven other goals and emphatically underscoring the prominence of these issues in global development and wellbeing.

The sub-Saharan Africa region has endured food insecurity since the mid-1980s (Maxwell and Smith, 1992). Africa is plagued by high levels of vulnerability to food insecurity as influenced by recurrent incidences of physical insecurity, drought, flood, ill-preparedness against shocks

[2] Eradicate extreme poverty and hunger by the year 2015.

and other risk factors. The Regional Strategic Analysis and Knowledge Support System (ReSAKSS), using "an economywide multimarket model", concludes that Africa "is exceptionally vulnerable" as a result of rising poverty rates, hunger, malnutrition and food dependency (ReSAKSS, 2008).The Intergovernmental Authority on Development (IGAD) region, where the Sudan is located, has probably experienced more episodes of hunger and food insecurity than the rest of the economic regions in Africa (IGAD, 2003). This comes as no surprise since the region is commonly characterized by a number of unfavourable geographical, political, social and economic conditions. Chief among these conditions are expanses of arid lowlands (80%), which receive an average 400 mm of rainfall per annum. Around 46 per cent of the land is described to be "unproductive", as farmland and forests account for only 25 per cent of the land (IGAD, 2003).

The Sudan, covering over a half of the IGAD region with an area of 2.5 million Km^2, has suffered chronic food insecurity and recurrent forms of transitory food insecurity worsened by a civil war that lasted over two decades (1983 – 2005). An ongoing civil war in the western part of the country (Darfur Region), which erupted in 2001, has only added to the problems. The semi-autonomous region of Southern Sudan, which acquired its status following the signing of the Comprehensive Peace Agreement (CPA) in January 2005, started experiencing severe food insecurity and humanitarian crisis since the war broke out. The United Nations World Food Programme (WFP) described Southern Sudan to be "the poorest and least developed region in Sudan and one of the poorest and least developed regions in the world" (WFP-VAM, 2007b). This portrayal says it all; vulnerability to food insecurity has stubbornly become characteristic of Southern Sudan. The 21-year war aggravated two worst episodes of hunger and starvation ever reported in the region(1984-85 and 1998), which claimed heavy death toll due to starvation, especially of women, children and the elderly (Bond, 1998). The direct cause of the crisis was drought, caused by El Niño[3] (HRW, 1999). One of the repercussions of these crises is the adverse effects on the agriculture and livestock sector, which before the war contributed

[3] "A warm current that flows along the equator from the date line and south off the coast of Ecuador at Christmas time" (WORDNET (2006) El nino. *WordNet.* Princeton, http://wordnet.princeton.edu.)

immensely (37% of GDP, agriculture alone) to the country's economy and the livelihoods of rural communities (Abbadi and Ahmed, 2006).

Against this bleak backdrop, close monitoring and evaluation of the situation would be helpful, if only there were reliable baseline and subsequent information. Another reason for finding good measures of food insecurity and the likelihood of hunger is in order to be able to monitor the indicators of the first Millennium Development Goal (MDG), that is, "eradicate extreme poverty and hunger". However, food security researchers and planners have encountered a mammoth task of finding valid, reliable and cost-effective measures for ascertaining the magnitude and scale of the risks involved (Maxwell, 1995). It is imperative that this limitation has dire implications in policy formulation and mitigating the problem of food insecurity. In as far as it is understood that the rank of the Millennium Development Goal number 1 must not have been an accident, finding better, pragmatic measurement criteria leading to versatile and robust indicators, is of critical necessity.

Measuring food insecurity at household level has even dogged food security analysts for many years. According to (Maxwell, 1995), two widely used methods for measuring household food security – household economy analysis and dietary diversity score – fail to yield convincing analysis of household food security. This study is based on analysis of possible causes of household food insecurity by examining a subset of data collected in a major household survey – the Sudan Household Health Survey (1996). Although the word "health" was central in the design of the survey (UNICEF mainly funded the activity), a separate questionnaire on food security was conducted, but the data were not formally analysed and the report did not cover information on food security (GoSS-MOH, 2008).

1.1 Importance of the study

The motivation for this Project stems from the conjecture that food security programmers and policy makers seem to have drawn a foregone conclusion that food security in Southern Sudan is a function of the war and displacement, combining with unfavourable environmental conditions such as drought and floods. This presumption is built on the premise that current interventions for mitigating food insecurity in the region seem to be inclined toward food aid and short-term relief rather than chronic food insecurity. It is also noted that what the USAID-funded Famine Early Warning System (FEWS NET) persistently outlines as causes of food insecurity are: decline in food crop production; poor harvest; food insufficiency; poor rainfall; drought; floods; large displacements and; food access problems (FEWS-NET, 2007).

Although some community and household data collection surveys or mini-surveys have been conducted over the years, they were done in an isolated fashion. Analysis of data from these surveys often either limited their findings to satisfying the purposes of the organisation involved or failed to present the national picture. FAO Sudan Information for Food Security in Action (FAO-SIFSIA, 2008) notes that some of the gaps and limitations facing the generation and use of food security information is reliance on out-of-date census and baseline data, lack of standard data collection and analysis duplication and lack of coordination among stakeholders. It may not be known for sure as to what has caused the lack of analysis and reporting of the food security dataset collected during the Sudan Household Health Survey of early 2006 (GoSS-MOH, 2008). It can only be perceived that the results of the survey might have been viewed as “overtaken by events”; that is, they would be too late for action, or the institutions that commissioned the survey lacked analytical capacity, or a combination of both.

Nevertheless, it is important to revisit the analysis of the readily available food security dataset to explore the use of robust statistical techniques in critically investigating household conditions, behaviours and attributes that could be associated with prevalence of food insecurity in Southern Sudan. It is anticipated that food security analysts could utilise similar data and their method of collection and utilise statistical analysis techniques for ascertaining the possible effect of household characteristics and livelihoods systems on food security. The analysis of livelihoods and food insecurity in Southern Sudan has been limited and unsupported by more robust statistical techniques and theory. It could be pertinent to find out if the presence of food insecurity in the households is a function of controllable conditions such as household size, availability of certain resources, and so forth. In statistics terminology, it is deemed important to investigate whether the *outcome variables* are explained by the identified *predictor variables* pertaining to the household. Alternatively, the aim of the study is to investigate the presence, magnitude, significance and efficiency of suspected variables that could predict the probability of occurrence of variables under study. Another value of the analysis could be to provide baseline information for subsequent future evaluations of the food security situation.

Further, the study attempts to identify suitable options for intervening in food security-related issues in Southern Sudan. Furthermore, the study is also expected to guide policy and strategy development for improving food security in Southern Sudan. Finally, the study aims at providing basis for further research into those variables determined to be statistically significant in their association with food insecurity outcomes.

1.2 Research problem statement

The underlying causes of household food insecurity in Southern Sudan have not been determined in scale and magnitude, using robust statistical techniques and hence the predicting food insecurity remained a dilemma.

1.3 Sub-problems

Methods and procedures used in the analysis employed in this Project are structured in an attempt to answer two pertinent questions. First, is there possible relationship between variables attributable to household composition, household endowments, entitlements and availability of main sources of livelihood on the one hand, and food consumption scores on the other? Alternatively, could food insecurity be a function of factors associated with, or perpetuated by, households; thus weakening their resilience to withstanding environmental, economic and physical security shocks? Secondly, do the statistical methods and techniques used in the analysis of the dataset lead to valid conclusions; thus vindicating their use in similar data collection surveys?

1.4 Hypotheses

The first study question above prompts testing of the null hypothesis (H_0) that there is no relationship between investigated household characteristics and food security indicators[4], and that any such relationship occurs by chance only. The null hypothesis is tested against an alternative hypothesis (H_1), which states otherwise, i.e. there is a relationship between the variables investigated and food security indicators. In statistics terminology, the null hypothesis states that each of the coefficients (β_j) of the explanatory variables (X_j) is equal to zero, that is, $\beta_1 = \beta_2 = \cdots = \beta_n = 0$; where $j = 1,2,\cdots,n$; $n =$ number of cases of the variable in the sample. The alternative hypothesis (H_1) is stated such that $\beta_1 \neq \beta_2 \neq \cdots \neq \beta_n \neq 0$. If the null hypothesis is rejected and the alternative hypothesis is accepted, it is then concluded that factors associated with the household could be responsible for household food insecurity. Otherwise, if we fail to reject the null hypothesis, it could be concluded that there is no association between the explanatory variables and the food security outcome under study; that is, if any relationship exists it could be due to chance alone.

1.5 Conceptual framework for describing the rationale of the study

The Project attempts to explore presence of a relationship between different explanatory variables (household characteristics and endowments) and outcome variables (food security index and incidence of household recovery from food security shock). In other words, the study attempts, using raw data from the Sudan Household Health Survey (2006) (GoSS, 2007), to take food security and household economy analysis to another level: to identify what could be the most influential determinants of household food insecurity in Southern Sudan. The conceptual framework of the study is presented diagrammatically in Figure 1.1:

---

[4] Food Security indicators treated as response or outcome variables are the Food Consumption Score (FCS) and Food Consumption Group (FCG).

Figure 1.1: Conceptual Framework of the Project

1.6 Study limitations

This study is based on analysis of secondary data. Moreover, two years have elapsed since data collection was carried out. Since then certain changes in population characteristics, mainly caused by return of internally displaced populations (IDPs) and refugees, might have taken place. Household characteristics could have also changed as people were relatively more settled at the time of data analysis than they were in 2006. Another limitation could be in the manner that the food security dataset was collected – as a side step in the structured interview of households – implying that cases of respondent fatigue might have resulted in missing data. Hence, this could lead to inefficiency of estimates (a case of large variability in responses).

1.7 Definition of terms

Food insecurity: “Lack of access to an adequate diet – which can be either temporary (transitory food insecurity) or continuous (chronic food insecurity”) (Bank, 1986)

Food Security: "Access by all people at all times to enough food for an active, healthy life" (Bank, 1986), or:

"A state that exists when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food which meets their dietary needs and food preferences for an active life" (FAO, 1996)

Hunger: "The uneasy or painful sensation caused by a lack of food. The recurrent and involuntary lack of access to food. Hunger may produce malnutrition over time.... Hunger ... is a potential, although not necessary, consequence of food insecurity" (Bickel et al., 2000).

Logistic Regression: A statistical technique of analysing data to determine the relationship between an outcome (dependent or response) variable and a set of independent (predictor or explanatory) variables; where the outcome variable is binary or dichotomous (Hosmer and Lemeshow, 2000).

Variable: In statistics, is a measurement unit that changes with respect to form, type, time, space or scale (own knowledge).

1.8 Study assumptions

The study generally assumes that changes in household characteristics have not changed in average terms. Further, in this text all statistical assumptions based on the logistic regression or linear regression modelling techniques hold.

1.9 Methodology

The Project will analyse secondary data from the Sudan Household Health Survey 2006. The Survey was modelled using the Multiple Cluster Survey (MICS) methodology (UNICEF, 2008).The raw dataset is in SPSS® (SPSS, 2008) format and analysis will be based on this computer package. Statistical log-linear models, such as the Logistic Regression model for binomial and ordered categorical data, will be employed to explore existence of relationships between response and explanatory variables. Statistical models will be aimed at tests of hypotheses. Data exploratory methods will also be used. Detailed description of the methodology is done in Chapter Four.

1.10 Dissemination of Results

The results of the Project will be presented to: (a) an audience of food security stakeholders in close consultation with the Africa Centre for Food Security and the University of KwaZulu-Natal, in form of seminars and publishing in ACFS web pages and; (b) Sudan Capacity Building Programme: Sudan Information for Food Security in Action (SIFSIA) stakeholder representatives; (c) relevant journals in form of articles; (d) Conferences in the broad area of food security.

1.11 Organisational structure of the mini-dissertation

The organisation of the main body of this Project is more as portrayed by the Conceptual Framework presented in Figure 1.1, is thought to be more lucid if presented as shown in Table 1.1.

Table 1.1 The organisation and rationale of the mini-dissertation

Chapter	**Title/Focus Area**	**Objective**	**Desired Outcome**
1	The problem and its setting	To portray a clear and broader picture of household food insecurity in the Sudan and to reinforce the importance of the study	Interest in the study attracted
2 and 3	Review of relevant literature	To present the context of household food insecurity and to describe the body of theory on statistical techniques used	Confidence in the methodology used for analysing, interpreting and concluding on findings instilled

Table 1.1. (continued...)

Chapter	Title/Focus Area	Objective	Desired Outcome
4	Methodology	To justify the validity and relevance of procedures used in sampling and data collection and explain the techniques used in data analysis	A deeper understanding of methodology used for arriving at valid findings, interpretations and conclusions inculcated
5	Results and discussions	To present the results of data analysis, interpret computer generated statistics and draw valid conclusions.	Acceptance of statistical techniques used and their application in similar studies.
6	Conclusion and recommendations	To give the synthesis of the broader picture of the study findings and establish what it has revealed in a nutshell. Recommend whether similar approach and methodology can be copied for further use.	Wide application of the statistical techniques used for modelling the data for establishment of relationships.

1.12 Summary

Food insecurity is a real menace in the world especially to the vulnerable and resource-poor populations in the least developed countries where Sudan is one. There are suggestions that the problem may persist, given current downward economic trends. Nevertheless, food security researchers and policy makers have often faced a dilemma in finding more robust measurement of food security risk and manifestation. Monitoring of food insecurity at household level on regular basis, is important and a necessity. However, this has not been adequately done to allow evidence-based and timely intervention. It is argued that national household surveys can provide an opportunity to measure the extent of house food insecurity vis-à-vis ascertaining the risk to household and individual vulnerability. Statistical techniques based on the logistic regression theory seem to offer the solution to the dilemma. The study applies logistic regression modelling for categorical data as discussed in the literature availed in the next Chapter.

CHAPTER TWO

REVIEW OF RELATED LITERATURE ON FOOD SECURITY MEASUREMENT APPROACHES

2.0 Introduction

This Chapter gives an account of available perspectives on measurement of food security at household level. The following Section specifically focuses on arguments made in different food security documentations, articles, monographs and books on food insecurity measurement methods adopted, options, challenges and recommendations. It is noteworthy that the literature reviewed use the terms “measurement of food security” and “measurement of food insecurity” interchangeably. This implies that a result of the presence or absence of a phenomenon, in measurement terms complement one another. That is, interpretation of results of either side of the dichotomy, point to the other side by deduction. However, for purposes of this study, focus is centred on measurement of food insecurity, as it is concerned with the problem (risk) not absence of the problem. An analogy is derived from the field of epidemiology which emphasises measurement of risk and not the absence of risk.

2.1 Review of existing approaches in measurement of food insecurity

2.1.1 The search for a “Gold Standard” for measuring household food insecurity

Lack of standardised measure of food insecurity has remained an issue of concern to many a food security analyst. Wolfe and Frongillo Jr. (2000) note that finding a method to best measure food insecurity has been “a subject of much debate”. They argue that this is partly due the difficulty related to defining food insecurity, which does not only include the well publicised composite components of lack of availability, access and utilisation but also perceptions about uncertainties of food insecurity.

When speaking of food insecurity what comes to mind quickly is malnutrition, hunger and absence of food in the house or in the local markets. This could be the main reason why measurement of food insecurity seems to be inclined towards these aspects. Indicators of malnutrition – as determined by anthropometric surveys, caloric intake and food consumption (dietary diversity), incidences of occurrence of hunger and food prices, feature highly in the literature on food insecurity measurement. Riely, et al. (1999) list some of the most commonly used indicators by food security programmes as: food production; income; total expenditure; food expenditure; share of expenditure on food caloric consumption and nutritional status. Maxwell (1995) believes that most food security analysts resort to measuring food consumption to escape the difficulty involved in making "complete analysis" of household food security; considering complexities surrounding household composition versus their resource-based income. Whereas this range of indicators directly provides information on the magnitude and the presence or absence of the problem, by themselves they leave out more of the information in concluding potential problem or an underlying risk. More so, such traditional way of measurement only tells one side of the story and fails to address possible effects of food insecurity. This could be the reason why existing food insecurity measurement methods and approaches are not planning-oriented but rather are tailored to monitoring and evaluation of programmes. Consequentially, the population is rendered unaware of a potential food insecurity risk awaiting them[5]. Review of literature in this study is therefore streamlined to probe what is documented to reinforce these arguments.

The significance of measuring *household* food insecurity (i.e. making the household as the unit of analysis) cannot be over-emphasized. It is important to recognise that the household is the optimal unit of data analysis (Bently and Pelto, 1991). Households are also referred to as "the social institutions through which individuals access food" (FANTA, 2003, Maxwell et al., 2003). Key among the reasons for treating a household as the unit of analysis is that it is a social unit. Although individual members of the household could have different characteristics and attributes relating to food security, households have many aspects that qualify them to be treated as units of analysis. Bickel, et al (2000) affirms that "food security is an essential, universal dimension of household and personal well-being" and that food security and hunger are "possible precursors to nutritional, health, and developmental problems". It is incontestable that wellbeing and hunger are attributes tying well with the individual person or the household that binds them and where they commonly share livelihoods entitlements and endowments. This therefore underpins the call for finding a

[5] This conclusion does not take into account a tool innovated by this scholar which is aimed at measuring household resilience using composite index for predicting food insecurity risk. In order to satisfy the curiosity of the reader, the indicator and measurement approaches are included in this dissertation.

measure that summarily concludes household food insecurity with high statistical accuracy and efficiency. Featuring prominently among the reasons that qualify a household as unit of analysis, is sampling based a sample frame obtainable from census offices. Furthermore, as is often defined in population and housing census training manuals, a household encompasses people who usually share food (meals), shelter and other livelihoods assets and has a head. Of overriding importance in our view, is the fact that the household readily yields quantitative data as opposed to other forms of data collection which are heavily dependent on qualitative approaches.

A number of food security experts have highlighted different purposes for measuring household food insecurity. Hoddinott (1999) sees the value of household-based measurement of food insecurity in terms of the need to “identify the food insecure” by assessing the severity of food insufficiency and to “characterize the nature of their insecurity”. Here focus is on the *intervention monitoring* purpose of household measurement. Riely, et al. (1999) underscore the importance of food security information which goes beyond programme monitoring and impact evaluation to “the design of relief and development interventions”. Human nutrition practitioners see the value of household food security measurement through the lens of their own domain. That is, to be able to monitor food utilisation at household level. Public health practitioners and leading international health organisations, such as UNICEF and WHO, are interested in information on mortality and morbidity especially of mothers and under-five children. Nandy, et al, (2003) construe a strong association between severe child malnutrition and ill-health.

Nevertheless, the complexity of food security, as a crosscutting discipline, has engrossed the challenge to finding a summative (or ‘gold standard’) measure of household food insecurity (Coates et al., 2003). Webb, et al. (2006) observe that a number of agencies lack a method for distinguishing households in terms of their food security levels so that they can target and evaluate their programmes reliably. FAO’s Hartwig de Haen, speaking at the closure of the International Scientific Symposium on Measurement and Assessment of Food Deprivation and Undernutrition points out that analysis of food insecurity still lack “a perfect single measure that captures all aspects of food insecurity” (FAO-FIVIMS, 2002). The USAID-funded Food and Nutrition Technical Assistance (FANTA) (2003) includes thirty-three indicators in its recommended list of indicators for measuring food insecurity access alone. Another difficulty to finding a breakthrough in measuring food insecurity is as observed by

Hendriks (2005, Gittelsohn et al., 1998) that the definition of food security is obscured by some terminology such as sufficiency and sustainability. Obviously the diversity of the dimensions and forms of food insecurity is a daunting limitation confounding to the measurement dilemma. Of importance, is to note that the authors cited above resoundingly see the need to finding simple and realistic measure of household food insecurity that can be labelled as “golden rule” combining rigour and statistical efficiency to conclude food insecurity from the household level upwards (Chung et al., 1997).

The study has outlined a number of food insecurity indicators described in a number of food security publications, featuring observations and findings of food security experts, academics and researchers. The search resulted in distribution of food insecurity indicators into five categories, namely; food sufficiency, food access, food utilisation, vulnerability and resilience to shocks and stresses. Table 2.1 summarises the available measures of household food insecurity as obtained in the literature reviewed.

Table 2.1 Measures and indicators of household food insecurity by category of information and source of data

Category of Information / Indicator	Measurement Method and Reference
Food sufficiency: – Household Food Insecurity Access Scale (HFIAS) – Depletion of stores (Maxwell, 1995)	– Survey of perceptions of experiences of lack of food (Coates et al., 2007) – Vulnerability assessments
Food access: – Household Dietary Diversity Score (HDDS) as a proxy measure – Household Food Insecurity Access Scale (HFIAS) – Months of inadequate household food provisioning – Household caloric and nutrient consumption – Expenditure gaps – Food gaps	– Calculation of food groups consumed over 24-hours’ period (Swindale and Ohri-Vachaspati, 2005) – Household caloric acquisition (Aliber and Modiselle, 2002, Hoddinott, 1999b, Swindale and Ohri-Vachaspati, 2005) – Food intake (Aliber and Modiselle, 2002, Hoddinott, 1999b) – Household Economy Analysis (SCF(UK), 2000) – Food Access Survey Tool (Coates et al., 2003)

Table 2.1 (continued)

Food utilization (nutrition and diet quality): – Anthropometry: wasting index; stunting; underweight; wasting – Adequacy of nutrition (number of eating occasions, dietary diversity and minimum daily caloric consumption)	– **Malnutrition surveillance (Gibson, 1990, WHO, 1995)** – **Household food consumption (Swindale and Ohri-Vachaspati, 2005)**
Vulnerability: – Coping Strategy Index – Food consumption (dietary energy supply; under-nutrition; etc...) – Health status (life expectancy and Under-5 mortality) – Nutritional status (Adult BMI; Under weight of under 5s; stunting; wasting)	– Vulnerability Analysis and Mapping (FIVIMS, 2002, WFP-VAM, 2008) – Strategies or ability to react to insufficient diet Index[6] (Maxwell et al., 2003, Maxwell, 1995)
Food Security Sustainability: – Status of livelihoods assets assessment (physical; financial; human; social; natural) – Household Economy Analysis – Mortality, nutritional status and food security in crisis situations – Household Resilience Index and Community Resilience Index (for Participatory Rural Appraisals)	– Sustainable Livelihoods Approach (Cahn, 2002, DFID, 1999, Frankenberger, 1992) – Household surveys; Key informant interviews; national socioeconomic surveys (DFID, 2001, Maxwell et al., 2003) – institutional and social network mapping (FSAU, 2006) – Food Economy Approach (Boudreau, 1998, SCF(UK), 2000) – The SMART Methodology (SMART, 2006) – Household resilience surveillance* (Lokosang, 2009)

* Unpublished work.

[6] These are short-term dietary changes (i.e. switch to cheaper food or eating less preferred food), eating less food per day (i.e. rationing), etc.

Table 2.1 gives enough evidence to confirm the diversity of measures of household food insecurity that differ in purpose, scope, scale, efficacy, type of data collection instrument, methodology and analysis approach. In fact, some of the measurements, such as those of nutrition, are based on data collected on individual members of the households and this adds to the complication. As already stated, the different purposes spelled out by different interventions targeting vulnerable households, could be the main reason for emergence of many indicators and different approaches to measurement of food insecurity at household level. It also seems that food insecurity researchers have focussed attention, in measurement of household food insecurity, on food access rather than the rest of the dimensions examined. Thus, vulnerability and food access dominate the literature of food insecurity measurement as underlined by the number and variety of indicators. The reason for this is well grounded, given the humanitarian and the economic nature of food insecurity and its implications. It is easy to see that some of the indicators, approaches and tools used in measurement are crosscutting in as far as the five dimensions of food insecurity are concerned. Finally, as the description of each measurement approach and indicator is not the focus of this study, the reader's attention is directed to the different bibliographies in the last column of Table 2.1. It is to be noted that only a few of the publications are cited here in a whole world of literature discussing measurement of food insecurity.

Five of the approaches or tools used in the measurement of food insecurity are listed in Table 2.2 and the main purpose of each approach is highlighted.

Table 2.2 Common household food insecurity measurement approaches, uses and main purposes.

	Approach	**Main Aim(s) or Focus**	**Main Uses**
1	Household Economy Analysis/ Approach (SCF(UK), 2000)	– Helps in operational decision making (Boudreau, 1998) – Quantitatively predicts that an event (e.g. crop failure or price change) is likely to affect people's ability to access food (SMART, 2006)	– Assessing relief needs, rationalising the use of food aid and early warning of food crises (Boudreau, 1998) – Understanding how poor people make ends meet and the reasons for rural-urban migration (Boudreau, 1998) – Developing policies against chronic hunger (Boudreau, 1998)

Table 2.2 (continued...)

2	Household Food Insecurity Access Scale	– Prevalence of household food insecurity access. – Monitoring of food insecurity access over time (Coates et al., 2007)	– Monitoring of food assistance programmes (Coates et al., 2007) – Assessment of programme impact (Coates et al., 2007)
3	Food Access Survey Tool (FAST)	– Food security-related programming and assessment for operational purposes (Coates et al., 2003)	– Guiding, monitoring and evaluating food security access operational interventions (Coates et al., 2003) – Assess poor people's perceptions of food insecurity and measure the experience of hunger (Coates et al., 2003)
4	Malnutrition Surveillance Measurement (Anthopometry)	– Highlights the need for a special food insecurity-related intervention[7] and the target population (Setboonsarng, 2005)	– Assessment of magnitude, distribution, and severity of a nutrition problem (Setboonsarng, 2005) – Proxy measure of a household's socio-economic level (Swindale and Bilinsky, 2006)
5	Dietary Consumption/ Food Intake	– Impact of household food access programmes – Dietary Diversity Score is a proxy measure of food access (Swindale and Bilinsky, 2006)	– Improved household food consumption (Swindale and Bilinsky, 2006) – Quality of diet (Setboonsarng, 2005) – A proxy for socio-economic level of the household (Swindale and Bilinsky, 2006)

The study has further reviewed and analysed seven widely used indicators highlighting their strengths and weaknesses as discussed by some authors of food security publications. Despite their dependence on qualitative data, the Coping Strategy Index, the Dietary Diversity Score and the Household Food Insecurity Score are rated as easy-to-use and readily calculable.

---

[7] Examples of malnutrition-targeting interventions can be found in MORRIS, S. (1999) Measuring nutritional dimensions of household food security. *Technical Guide No. 5.* Washington D.C., International Food Policy Research Institute..

Table 2.3 Strength and weaknesses of seven major indicators used in measuring and monitoring household food insecurity

	Indicator	Strengths	Weaknesses
1	Coping Strategy Index (CSI)	– "Rapid measure of short-term household food insecurity" (Maxwell et al., 2003) – Easy to implement and directly captures perceptions of availability and vulnerability (Hoddinott, 1999a) – "Good proxies for food intake; etc." (Christiansen and Boisvert, 2000, Hendriks, 2005, Maxwell et al., 2003) – "Ability to identify changes in household conditions as a result of emergency food aid operations" (Hendriks, 2005)	– The assessment cannot be repeated for the same community, as respondents may alter their responses to the coping strategy behaviour questions in subsequent rounds (Hendriks, 2005). – Caution needs to be taken in interpretation of results as some coping strategies are reversible while others are irreversible (Gillespie et al., 2001, LOEVINSOHN and Gillespie, 2003) – High susceptibility to misreporting (Hoddinott, 1999a)
2	Dietary Diversity Score (DDS)	– Easy to use and straightforward taking less than 10 minutes to complete a questionnaire (Hoddinott, 1999a, Swindale and Bilinsky, 2006) – Tracks seasonal changes in food security (Hoddinott, 1999a) – Enables examination of food insecurity at the household and intra-household levels (Swindale and Bilinsky, 2006)	– If responses are not weighted, the method does not record quantities (Hoddinott, 1999) – It is not possible to estimate the extent to which diets are inadequate in terms of caloric availability, unless the frequency of consumption of particular diets is probed (Hoddinott, 1999a)
3	Child Malnutrition	– Relatively easy data collection (Nandy et al., 2003) – Based on quantitative data and therefore more objective – Can lead to more summary descriptive statistics. – Powerful for evaluation of household, sub-national and national food insecurity status and programmes	– Require a large sample to arrive at efficient statistics for concluding the nutritional level of an area. – Results not easy to interpret *at any one time* i.e. requires time series (Setboonsarng, 2005)

Table 2.3 (continued...)

	Indicator	**Strengths**	**Weaknesses**
4	Household Food Insecurity Access Scale (HFIAS) Anthropometry	– Method has been tested and validated in some developing countries and generated required indicators (Coates et al., 2003, Frongillo and Nanama, 2006, Webb and Rogers, 2003) – Proven sensitivity to different cultural contexts (Coates et al., 2007)	– Heavily depends on individual's perceptions of food security access aspects and thus cannot be standardised with regard to different cultural contexts – requires adaptations to local settings and compromising comparability of information across countries or regions
5	Dietary Energy Consumption (household caloric consumption	– Can produce more accurate measures of individual caloric intake if measured correctly (Hoddinott, 1999b) – Can indicate within household disparities of food insecurity status (Hoddinott, 1999b)	– "Too cumbersome" to be used for targeting food aid (Chung et al., 1997) – Questionable reliability of data sources (Boudreau, 1998) – Reliance on "expert" data collectors and analysts (Hoddinott, 1999b) – Require repeated measurements (Hoddinott, 1999b)
6	Household Economy Analysis	– Provides decision makers with an understanding of the picture of a rural economy; thus helping in food aid programming (Boudreau, 1998)	– Reliance on "expert" analysis (own observation) – "Requires a high degree of training, well educated, committed and enthusiastic staff (Boudreau, 1998) – Non-verifiability of results, costly, time consuming and impracticable (Boudreau, 1998)
7	Sustainability Assessment based resilience	– Quick, does not require rigorous calculation and predictive of impending vulnerability – Can be used for surveillance of food insecurity - Instantaneous results; readily shared with the respondent and easily summarised	– Untested. Therefore, no universal agreement on method. – Weighting of scores is arbitrary i.e. susceptible to prolonged debate. – How to adjust scores on the percentile scale threshold could be challenging

Note: Where no citation is provided, the postulations are those of the author

It seems that indicators measuring food access which are based on quantitative data, although they enable predictive analysis of household food insecurity (i.e. they permit statistical inferences), they require “expert” involvement, rigorous and time consuming data collection and analysis. This requirement does not augur well with the nature of food insecurity as a crisis situation requiring timely decision making and immediate intervention. This could be the main reason why food security programmers and decision makers are often faced with the bunch of challenges outlined above.

Hoddinott (1999a) paints a clearer picture of the advantages and disadvantages of four of the above indicators in Table 2.3 in terms of certain intrinsic qualities.

Table 2.4 Comparison of methods for monitoring household food insecurity

	Food Intake	**Household Caloric Acquisition**	**Dietary Diversity**	**Coping Strategies**
Data collection costs	High	Moderate	Low	Low
Time required for analysis	High	Moderate	Low	Low
Skill level required	High	Moderately High	Moderately low	Low
Susceptibility to misreporting	Low	Moderate	Low	High

Source: Hoddinot (1999a)

Table 2.4 reveals that the Dietary Diversity Score and the Coping Strategy Index both possess relatively better strength for application in rapid household surveys. An important consideration of the approaches aimed at generating the two indicators is to undertake repeated surveys on the same population in order to be able to monitor the risk of food insecurity.

2.1.2 Approaches for measurement of the determinants of household food insecurity

The importance of understanding the determinants of household food insecurity or food insecurity in general, is paramount. As Chopak (2000) puts it, one reason is to help in the design of appropriate food insecurity mitigation actions. Hoddinot (1999b) conceptualises the determinants of chronic food insecurity from a two tier perspective: endogenous determinants and exogenous determinants, by mirroring them in a conceptual framework of Figure 2.1.

The encircled numbers in the framework indicate seven categories of possible determinants of household food insecurity. The framework demonstrates that household food insecurity is engulfed by three environmental factors which are referred to here as exogenous determinants. These are: physical, government policies and social determinants. The group of endogenous determinants include: (i) household resources or endowments in form of labour and capital; (ii) the activities that the households are involved in, or that they allocate their endowments to, and these include food production, cash crop production, non-agricultural income-generating activities and private/public transfers ; (iii) market prices; (iv) consumption options; (v) household healthcare and health environment and finally; (vi) feedback effects of the food in/security level; that is, the benefits spilling over from a rise in income and improvement (or degeneration) in vulnerability that increase (or deplete) the resource base of the household. Food and Agricultural Organisation (2000) describes the causes of household food insecurity vulnerability to be "difficult to measure". It further categorises the causes of food insecurity as: (a) availability of quantity and quality of household food; (b) physical and economic access to food. Riely, et al (1999) adds the dimension of food utilisation to the equation. These authors reinforce views by others stating that "adequate food availability at the aggregate level is a necessary, although not sufficient, condition to achieve adequate food access at the household level, which in turn, is necessary but not sufficient for adequate food utilization at the individual level". This chain of relationship of the three dimensions of food in/security is represented in Figure 2.2.

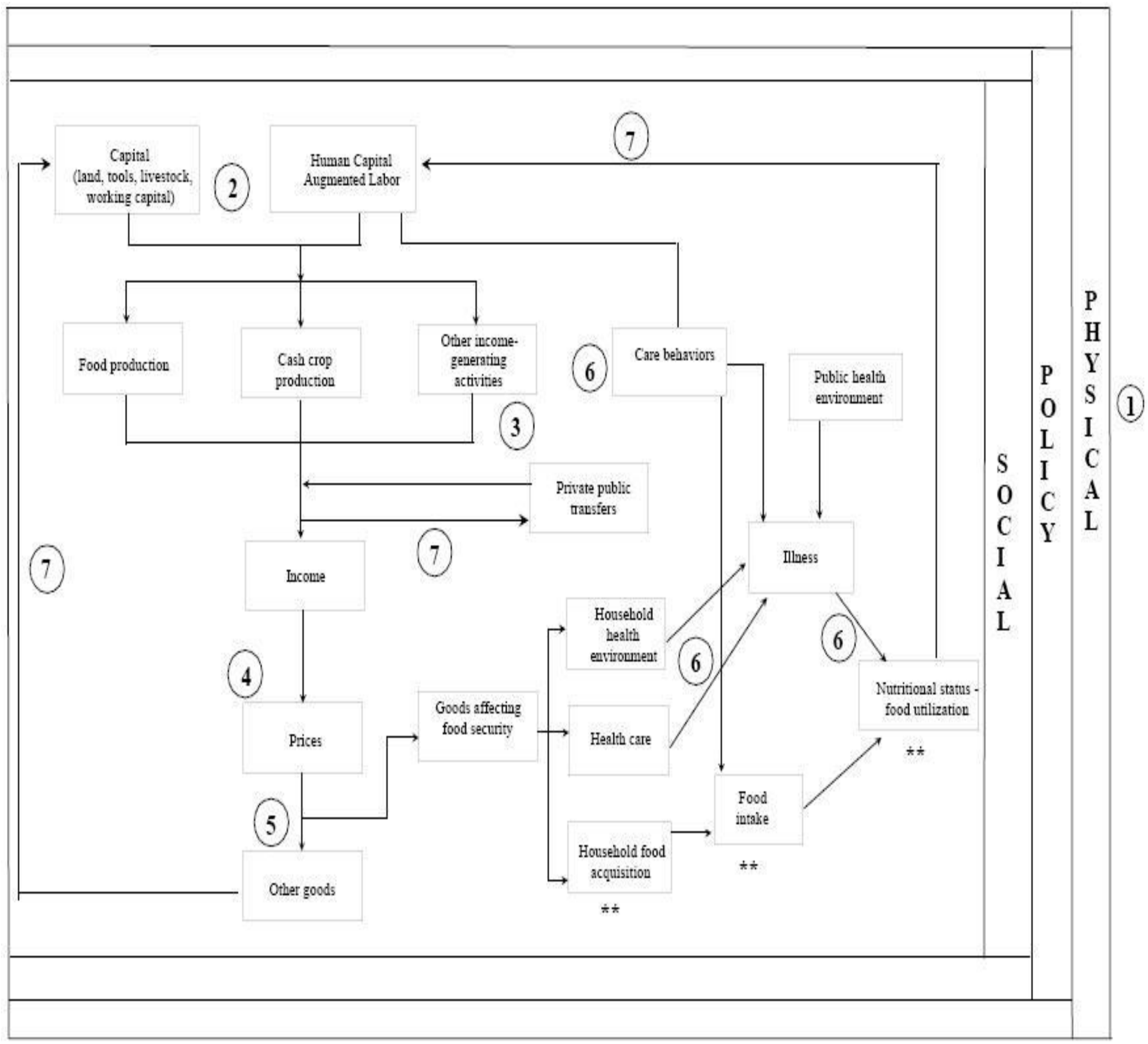

Figure 2.1 Determinants of household food (in)security (Hoddinott, 2000)

Riely, et al (1999) suggests that the causes of food insecurity could be understood in the context of the food security framework. The diagram points out that food availability is influenced by stocks available locally, imports, food aid and local food production. Food production is influenced by cash income which in turn is influenced by the household's resources and underlying influences. By reciprocation, food availability influences food prices which in turn influence market purchase of food. Food access is affected directly by food production, the market and cash transfers (i.e. government/private remittances or kinship support), and indirectly by food availability through food production prices.

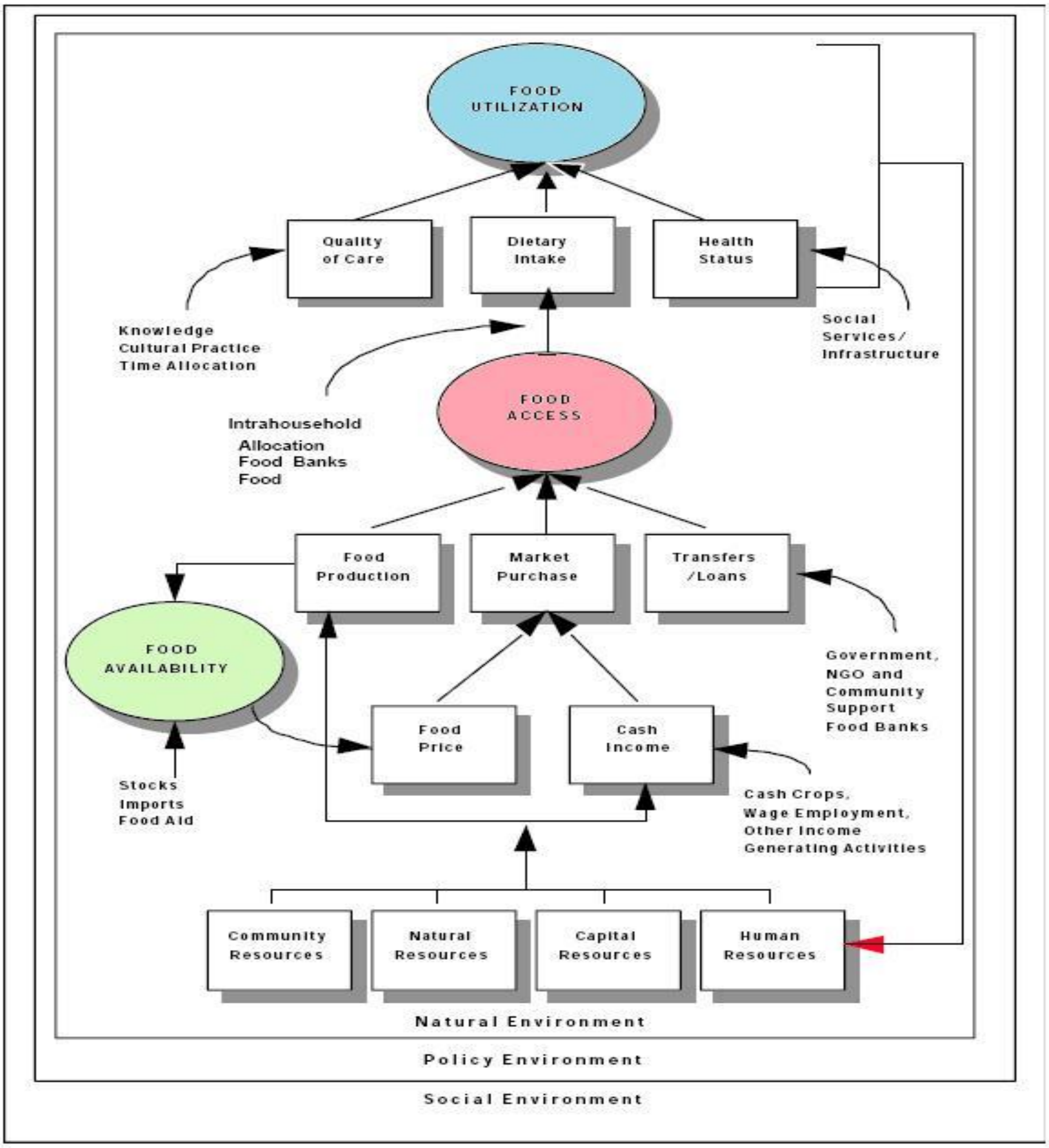

Figure 2.2: Food security conceptual framework (Riely et al., 1999)

Understanding the causes of food insecurity conceptualised in this way, is easy to grasp. However, the real challenge comes when attempting to know the magnitude, degree and the relative importance of the causes as well as how to obtain the sort of data that may lead to conclusion of a relationship, statistically speaking. Riely, et al (1999, p. 19) recommends that understanding of the causes of food insecurity can be bolstered by routinely collected data by ministries of Agriculture, Health or Planning or by Demographic and Health Surveys (DHS) or secondary data from existing studies and reports. However, the post-conflict setting of Southern Sudan underlines the problem of lack of structures set up for routine data collection by government institutions let alone acute technical capacity gaps.

CHAPTER THREE

REVIEW OF RELATED LITERATURE ON STATISTICAL THEORY USED

3.0 Introduction

This Chapter entails an explanation of the concepts leading to the choice of the statistical technique used and describes statistical approaches employed in the analysis of data with the dependent variable being on categorical and binary scale. It is to be noted that the nature of the dataset in the study dictates the choice of the technique and approaches. In this dataset the response variable under-study, Food Consumption Score (FCS), takes the form of ordered categorical (or ranked) responses. This variable is calculated based on scoring of groups of food eaten in the past week of the survey and the frequency or number of times of eating the food during the week. It is then transformed from a continuous scale variable to an ordinal categorical variable with three categories namely: *poor*, *borderline* and *good* (see Chapter 4). However, the variable may also be transformed into a dichotomous or binary variable. In this Chapter, the theory surrounding a response variable of the 'binary' type will be reviewed first. Then the statistical theory on ordered categorical data of the response variable FCS follows.

3.1 The binomial distribution and the standard normal distribution

A response variable is stated as 'binary' if any of its observations takes one of two possible forms. An outcome may occur or it may fail to occur. In the case of the incidence of recovery from shock, a household either failed to recover from some food security incidence of shock – in which case food insecurity was experienced – or it did recover. That is, it did not experience food insecurity. The form of distribution of these observations is said to have the **Bernoulli** distribution, which is a special case of a **binomial** distribution (Collet, 1991).

3.1.1 The probability distribution of a 'success' response

As sated, a binary response is either a *success* or a *failure*. If interest is centred on success, the probability of success, p, becomes important. Since the response is unknown before collecting the data, it can take a quantity known as random variable, denoted by R, which may take one of two values; *0* or *1*. Therefore, the probability that *R=1*, expressed as *P(R=1)* (or probability of *success*) can be given as *p*. That is, *P(R=1) = p*. It follows that the probability of failure, i.e., *P(R=0) = 1-p*. If *r* is the observed value of the random variable, with a possible value of *0* or *1*, the probability distribution of the response R is $P(R = r) = p^r(1-p)^{1-r}, r = 0{,}1$, which is known to be the Bernoulli Distribution (Collet, 1991).

In can be shown mathematically that the mean of *R* is $E(R) = 0\times P(R=0) + 1\times P(R=1) = p$ and that the variance of *R*, denoted $Var(R)$, is $E(R^2)-[E(R)]^2 = p(1-p)$. The interpretation of these parameters is that the distribution of binary responses has a constant mean or the same probability and variance (the spread of observations) *p(1-p).*

Where the response variable is suspected to be determined by some condition or a set of conditions, also known as explanatory variable(s), the condition(s) cause a situation where occurrences of the response are grouped, i.e., a sequence of repeated binary responses: successes and failures occur per category of the given condition. For example, a number of households that do not have farms or home gardens, that are female-headed and that did not own livestock might have shown manifestations of incidences of food insecurity as well as no incidences of food insecurity.

Let us suppose that in each group of observations there are *n* responses and that the random variable associated with success or failure is R_j for $j=1, 2,..., n$. As the response variable has the same probability of occurrence, each of the random variables $R_1, R_2,...., R_n$ will have a Bernoulli distribution in the form of $P(R_j = r_j) = p^{r_j}(1-p)^{1-r_j}$, where r_j is the observed binary response for the j^{th} response within a group, taking the value *0* or *1* for $j=1, 2, ..., n$. Letting y = total number of successes observed in a group of *n* responses, under the assumption that each response occurs independently within the group, it results that *y* is an observed value of a random variable Y associated with total number of successes out of *n*. Then from the domain of random variables, $R_1, R_2, \cdots, R_n$, $y = r_1 + r_2 + \cdots + r_n$, the sum of binary observations, is the observed value of the random variable $Y = R_1 + R_2 + \cdots + R_n$ (Collet, 1991).

In a sequence of *n* binary observations containing *y* successes and *n-y* failures, the probability of successes is given by

$$P(Y = y) = \binom{n}{y} p^y\ (1-p)^{n-y}\ ; y = 0,1,\cdots,n$$

as shown by Collet (1991: p. 19). This is a binomial distribution in the random variable *Y*. That is *Y* is binomially distributed with parameters *n* and *p* (also stated *Y~B(n;p)*). We need to recall that the *n* binary responses are assumed to be independent of one another. An important property of this distribution is that the mean (or expected value) of the random variable *Y* is *np* and the variance is *np(1-p)*, again as shown by Collet (1991, p.19), who also derives the standard *normal distribution* of random variable *Z* defined by

$$Z = \frac{Y - np}{\sqrt{[np(1-p)]}} \qquad (3.1)$$

which is the approximation of the binomial distribution to a normal distribution, as *n* tends to be really large, with mean zero and variance equals to one.

3.1.2 Methods of inference based on the 'success' probability

Collet (1991) has extensively discussed inference about the success probability statistic. Note that statistical inference is concerned with three issues. First, the *precision of an estimate*, i.e., how *precise* or how *reliable* is the sample statistic observed so as to predict the population parameter? Second, how *confident* are we about the estimate? Third, what is the *significance* of the difference between the observed statistic and a preconceived parameter or measure of association between the variables?

In answering the *first* question, we show how the parameter estimate (denoted as $\hat{p}$) of the success probability, *p*, is obtained. In principal it is conceived that as the number of observations, *n*, become really large, the estimate of the success probability, $\widehat{p}$, tends to become more reliable. This reliability is then ascertained if we know the *standard error* (or the *standard deviation*) of the estimate. The binomial distribution of the random variable *Y*, denoted as *B(n,p)*, suggests that the variance of an observation y is *np(1-p)*. It follows that the *variance* of the estimate of $p, \widehat{p}$, is

$$Var(\hat{p} = \frac{y}{n}) = \frac{p(1-p)}{n}$$

which can be estimated as

$$\widehat{var}(\hat{p}) = \hat{p}(1-\hat{p})/n$$

Obtaining the square root of this expression leads to the standard error (denoted as *s.e.*) of $\hat{p}$

$$s.e.(\hat{p}) = \sqrt{\left(\frac{\hat{p}(1-\hat{p})}{n}\right)}$$

If the standard error of the estimate $\hat{p}$ relative to the sample size n is large, it can be concluded that the estimate is *less precise*. Otherwise if the value of *s.e.* is small, the estimate is good and statistically *more reliabl*e.

Secondly, we derive the *confidence interval* regarding the true success probability *p*. That is, the range of values around the estimate where we expect the "true" population probability of success, *p*, to be located with a given *level of confidence* or certainty (StatSoft, 2007). The set level of confidence is also described by Collet (1991) as "the probability of including *p*" and usually given as *0.90, 0.95 or 0.99.* Setting the level of confidence as $1-\alpha$, where α is a small positive value gives a confidence interval of *100(1-α)%* for *p*. The value of α is usually 0.10, 0.05 or 0.01.This means if the sample selection were to be repeated a number of times, there would be a *100(1-α)%* chance of the interval containing the true probability of success. Being an interval of values, it has a lower limit (denoted p_L) and an upper limit (dented p_U), which are the smallest and largest binomial probability of success, respectively. Either probability, which equals to the observed proportion, y/n, has a probability that is at least $\alpha/2$ (Collet, 1991; p. 23). These probabilities are given by the expressions

$$\sum_{j=y}^{n} \binom{n}{j} p_L^j (1-p_L)^{n-j} = \frac{\alpha}{2} \qquad (3.2)$$

and

$$\sum_{j=0}^{n} \binom{n}{j} p_U^j (1-p_U)^{n-j} = \frac{\alpha}{2} \qquad (3.3)$$

which compute the lower and upper limits of the probability that a binomial random variable with parameters (n, p_L) and (n, p_U) respectively take the value of *y* or more and the value of *0* or more. The values of p_L and p_U are readily derived from Tables of Binomial Distribution for given values α, *y* and *n* , which are usually appended to many statistics text books (Collet, 1991; p. 23).

Note that as the confidence interval (p_L, p_U) depends on the sample size, *n*, and the variance (the spread of observations about their mean), the larger the sample size, the more reliable is its calculated probability value and vice versa. Note also that the calculation of confidence interval is based on the assumption that the variable is normally distributed in the population. Hence unless the sample size is large enough the estimate may not be valid if the assumption of normality is not met.

The *normal approximation* to the binomial distribution, constructed based on the percentage points of the standard normal distribution[8], provides the real answer to deriving the confidence interval for the 'true' success probability. Explained pictorially by Figure 3.1, if the random variable Z has a standard normal distribution, the upper (*100α/2*)% point of the distribution is $z_{\alpha/2}$, expressed as $P(Z \geq z_{\alpha/2} = \alpha/2)$. This probability density area is the shaded area to the right of $z_{\alpha/2}$. By symmetry, the lower (*100α/2*)% point of the binomial distribution is equal to $-z_{\alpha/2}$ such that $P(Z \geq -z_{\alpha/2} = \alpha/2)$. This probability distribution of the lower end of the random variable *Z*, is the shaded area to the left of the value $-z_{\alpha/2}$ in Figure 3.1. It follows that $P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}) = 1 - \alpha$. Statistical tables featuring the percentage points of a standard normal distribution, which are commonly found at the back of a number of statistics textbooks, can provide the values of $z_{\alpha/2}$ for a given level of α. For example, we can obtain the values of $z_{\alpha/2}$ for α*=0.1, 0.05* and *0.01*which are equal to *1.645, 1.960* and *2.576*, respectively (Collet, 1991).

[8] The standard normal distribution is a normal distribution with mean 0 (μ*=0*) and standard deviation 1 (σ^2*=1*).

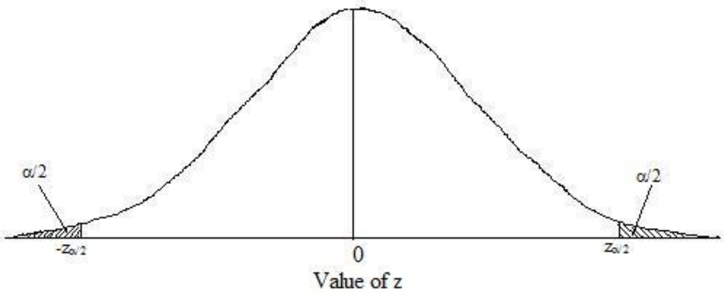

Figure 3.1: The upper and lower α/2 points of the standard normal distribution (Collet, 1991).

Now, re-expressing equation 3.1 by dividing the numerator and the denominator by *n* (i.e. multiplying by *1/n*) and replacing *y*, the observed value of *Y*, with $\hat{p}$ we get

$$\frac{\hat{p}-p}{\sqrt{[p(1-p)/n]}} \tag{3.4}$$

which has an approximate standard normal distribution of the area under the curve (Figure 3.1) between $-z_{\alpha/2}$ and $z_{\alpha/2}$, or that

$$P\left\{z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{[p(1-p)/n]}} \leq -z_{\alpha/2}\right\} = 1-\alpha \tag{3.5}$$

Replacing *p(1-p)* with $\hat{p}(1-\hat{p})$, that is, further approximating the value of the success probability, leads to equation 3.5 becoming

$$P\left\{-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{\hat{p}[(1-\hat{p})/n]}} \leq z_{\alpha/2}\right\} = 1-\alpha. \tag{3.6}$$

Since the expression $\sqrt{[(\hat{p}(1-\hat{p})/n]}$ is the standard error *s.e.* of $\hat{p}$, equation 3.6 becomes

$$P\left\{-z_{\alpha/2} \leq \frac{\hat{p}-p}{s.e.(\hat{p})} \leq z_{\alpha/2}\right\} = 1-\alpha. \tag{3.7}$$

It therefore follows that

$$\hat{p} - z_{\alpha/2}s.e.(\hat{p}) \leq p \leq \hat{p} + z_{\frac{\alpha}{2}}s.e.(\hat{p}) = 1-\alpha, \tag{3.8}$$

thus giving the required lower limit and upper limit of the confidence interval of the 'true' success probability, *p*. In interpreting the result, we say that if *p* is normally distributed with mean zero and unit variance, i.e. N~(0,1), we can be (100α)% confident that its true value lies between $\hat{p} - z_{\frac{\alpha}{2}} s.e.(\hat{p})$ and $\hat{p} + z_{\frac{\alpha}{2}} s.e.(\hat{p})$.

The third measure of inference about the estimate is to test the hypothesis that the value of p is equal to some pre-specified value, say, p_0 and to estimate the significance of any difference between the two values. In other words, we need to provide statistical evidence to suggest that either the 'true' (population) success probability differ significantly, or there is no adequate basis to conclude there is significant difference. That is, if there is any difference it might have occurred by chance only. A 'conservative' hypothesis commonly known as the '*null hypothesis'* (denoted H_0), would be stated as H_0: $p=p_0$. A 'complementary' hypothesis, widely termed the '*alternative hypothesis*', would be expressed to state there is a difference between the 'true' success probability and the theoretical probability, p_0. This is given as H_1:$p<p_0$. Rejection of the null hypothesis is suggested if the observed probability of success is too small. The null hypothesis formulated in this way means a one-tailed test which means the significance level, α, is not to be divided by *2* as p is distributed between *0* and y as in

$$\sum_{j=0}^{y} \binom{n}{j} p_0^j (1 - p_0)^{n-1}$$

The relative size of this probability determines the conclusion regarding rejecting or not rejecting the null hypothesis H_0. A reasonably large probability indicates a high number of successes relative to the number of trials n and that leads us not to have enough grounds to reject the null hypothesis that $p = p_0$. In contrast, a relatively small probability indicates small number of successes, an outcome which favours the alternative hypothesis H_1 which is that $p < p_0$ and leading to rejection of the null hypothesis H_0. In the second case, the probability of y successes, $P(Y=y)$, is smaller than α, the significance level of the hypothesis test and we conclude that the difference between $\hat{p}$ and its hypothesised value, p_0, is significant at the $100\alpha\%$ level of significance (Collet, 1991).

Quite often the output of statistics software presents the actual probability of success otherwise commonly known as the *p-value*, to warrant rejection or 'acceptance' of the null hypothesis. Collet (1991) describes the *p*-value as "a summary measure of the weight of evidence against H_0". The *p*-value is the *probability of error* involved in accepting the observed result as valid or representative of the population (StatSoft, 2007). A 'rule of thumb' for rejecting H_0 usually provides for rejection in situations when the *p*-value is given as in Table 3.1 below.

Table 3.1 The interpretation of the *p*-value and conclusion of significance regarding H_0

	***p*-value Range**	**Interpretation of level of evidence against H_0**	**Significance of *p*-value**	**Decision**
1	$\boldsymbol{p} \geq 0.1$	There is *no evidence* against H_0	Not significant	Do no reject H_0
2	0.05< $p \leq 0.1$	There is *slight* evidence against H_0	Borderline significant	Reject H_0
3	0.01< $p \leq 0.05$	There is *moderate* evidence	Somewhat significant	Reject H_0
4	0.001< $p \leq 0.01$	There is *strong* evidence	Highly significant	Reject H_0
5	$\boldsymbol{p \leq 0.001}$	There is *overwhelming* evidence	Very highly significant	Reject H_0

The judgements given above are those of a one-tailed or one-sided hypothesis test as we know that the alternative hypothesis H_1 states that $p < p_0$. Otherwise, if H_1 had stated that $p \neq p_0$, testing would have been based on a two-sided test, as we could aim at testing $H_1: p < p_0$ and $H_1: p > p_0$. Collet (1991, p.27), states that using one-tailed test is "somewhat controversial" and hence recommends using a two-tailed test.

3.1.3 Logistic Regression Model for Binary Data

Logistic Regression is in the category of statistical models known as *Generalised Linear Models* (Collett, 1991, p.56). Logistic Regression allows one to predict a discrete outcome from a set of variables that may be continuous, discrete, dichotomous, or a mix of any of these. Generally, the dependent or response variable is dichotomous, such as presence/absence or success/failure. Thus, in instances where the independent variables are a categorical, or a mix of continuous and categorical, Logistic Regression is preferred. According to Hosmer and Lemeshow (2000) Logistic Regression, like other statistical modelling techniques, has the goal of finding the best fitting model to describe the relationship between two variables.

There are two types of Logistic Regression models for binary data which can be fitted to the data. The first type of these models is the S***imple Logistic Regression model***, which involves modelling a relationship between one explanatory variable and the binary response variable.

The second type is the ***Multiple Logistic Regression Model***, which models *k* explanatory variables each with *m* levels. In both types of models, SPSS[9] LOGISTIC REGRESSION or SAS PROC LOGISTIC procedures are used for analysing the dichotomised data (SPSS, 2008). Estimates of *odds ratios* and *chi-square* statistics are calculated based on the *Maximum Likelihood* (Collet, 1991, Cox and Snell, 1989, Menard, 1995). The principles and theory guiding modelling of binary data using a Logistic Regression Model will first be discussed. A summary of parameter estimates is given for measuring relationships between a selection of potential explanatory variables and the binary outcome variable. Finally, model examination and selection is performed for the Multiple Logistic Regression method (Collet, 1991, Hosmer and Lemeshow, 2000, Menard, 1995).

3.1.4 The response data and preliminary modelling concepts

A model for binary data may be best understood when presented from the point of view of a relationship between two variables presented using a 2×2 contingency table and constructing inferential statistics for estimating and testing a relationship. Table 3.2 shows a relationship between a binary response variable with two levels and an explanatory variable, also with two levels, *0* and *1* indicating a “No” and “Yes” categories of the variable. A case in point is to explore a relationship between a household’s food insecurity status and the length of stay in the locality where it was interviewed. Food security status would have two levels, say, food insecure (S) and generally food secure (F). The length of stay would be categorised as *0 = stayed for less than 1 year* and *1 = stayed for over one year*. Food insecurity status would be the dependent or response variable and length of stay would become the independent or explanatory variable. Since there are two levels of either variable, a 2×2 contingency table is constructed.

[9] Statistical analysis software developed and marketed by SPSS, Inc.

Table 3.2 A 2×2 contingency table for binary data

	Failures	Successes	Total	Proportion of successes
Group 0	$n_0 - R_0$	R_0	n_0	R_0/n_0
Group 1	$n_1 - R_1$	R_1	n_1	R_1/n_1
Total	$n_0 + n_1 - R_0 - R_1$	$R_0 + R_1$	$n_0 + n_1$	

Source: Cox and Snell (1989)

In Table 3.2 the rows represent an independent variable composed of two groups of households, *0* and *1*, of sizes n_0 and n_1, respectively. The columns represent a binary dependent response or outcome variable. The values R_0 and R_1 are random numbers of the success of group *0* and the success of group *1*, respectively (Cox and Snell, 1989). Assuming that all households respond independently with probability of success depending only on the group, the probability of observing a success of household in group *0* can be expressed as $P(R_0)=\phi_0$ and the probability of a household responding with a "success" in group *1* is $P(R_1)=\phi_1$. However, Logistic Regression is concerned with probability of observing a success for an individual household in each of the two groups. This is expressed as a proportion of successes in group *0* and then in group *1*. (Cox and Snell, 1989 and Collett, 1991, p. 56).

Suitable statistics for measuring a relationship between two variables expressed in the form of Table 3.2 are those of the *odds ratio* and the *chi-square* (χ^2). The odds ratio statistic is used for measuring the magnitude of the relationship, while the chi-square statistic is used for testing an association between two variables (Hosmer & Lemeshow, 2000). There are often two common methods for calculating the chi-square statistic: the *Pearson's chi-square* or through two *Maximum Likelihood* statistics called the *Efficient Score* (*Z*) and *Fisher's Information* (*V*). In the latter case,

$$\chi_1^2 \cong \frac{Z^2}{V}$$

The efficient score and Fisher's information also give an estimate of the *log-odds ratio* as

$$\hat{\theta} = \frac{Z}{V}.$$

A maximum likelihoods derivation of the Efficient Score and Fisher's Information is found in Collet (1991) on pages 342-3 (Appendix B). The statistics software that will be used readily gives these statistics for testing the association between explanatory variables and response variables.

The odds ratio

Prior to defining the *odds ratio*, we need to define what is meant by "*odds*". We say, "the odds of a success", which is defined as the ratio of the probability of "success" over the probability of "failure" (Collet, 1991). From Table 3.2, the odds of success in group *0* are $\emptyset_0/(1-\emptyset_0)$, and the odds of success in group 1 are $\emptyset_1/(1-\emptyset_1)$. In comparing the two groups, we say that the odds of a success in group 1 relative to group 0 are:

$$\psi = \frac{\emptyset_1(1-\emptyset_1)}{\emptyset_0(1-\emptyset_0}.$$

In interpreting this statistic, when ψ > 1, we say the odds of success favour group 1 than group 0 and when ψ < 1, we say the odds of observing a success are more in group 0 than in group 1. Hence, the odds ratio is the measure of the difference between two success probabilities related to two comparable groups (Collet, 1999). It follows that the estimate of the odds ratio is given by

$$\hat{\psi} = \frac{\hat{\emptyset}_1\left(1-\hat{\emptyset}_1\right)}{\hat{\emptyset}_0(1-\hat{\emptyset}_0)}$$

and we say the odds of a success are ψ times more (or less) in group 1 than in group 0. If $\hat{\psi} \approx 1$, (i.e. ψ is very close to 1), it can be interpreted that there is no change in odds between the two groups, suggesting there is no association between group of households and the response/outcome variable. Evidently if $\hat{\psi} > 1$, the odds of success in group 1 relative to group 0 are more and similarly, if $\hat{\psi} < 1$, the odds are less in group 1 relative to group 0 (i.e. they are more in group 0). In both cases of the inequality to unity, there is suggestion of an association between the explanatory variable and the response variable.

The value of ψ tends to be normal on the logarithmic scale for a large sample. Hence, when the individual or a household's responses are not known, the odds ratio is better defined in terms of *log-odds ratio* (Collet, 1991). If we denote the odds ratio by θ, then

$$\hat{\theta} = \ln(\hat{\psi}) = \ln\left\{\frac{\emptyset_1 - (1-\emptyset_0)}{\emptyset_0 - (1-\emptyset_1)}\right\}$$

The standard error of the log-odds ratio, in terms of Table 3.2, is that shown by Woolfe (1955) and Schlesselman (1982) as

$$se(\ln\hat{\psi}) = \sqrt{\frac{1}{R_1} + \frac{1}{(n_1 - R_1)} + \frac{1}{R_0} + \frac{1}{(n_0 - R_0)}}$$

The 95% confidence interval for $\ln\hat{\psi}$ is $\ln\hat{\psi} \pm 1.96se(\ln\hat{\psi})$. Then the 95% confidence interval (*CI*) for $\hat{\psi}$ is obtained by taking the exponents of the lower and upper limits. It is simpler to calculate the estimate of the odds ratio and the chi-square statistic, from values of the efficient score and Fisher's information as obtained in the following Sub-sections.

Pearson's chi-square statistic:

If we denote the number of successes in group *j* by y_j, and the number of binary observations in the group by n_j, it follows that the number of failures is $n_j - y_j$, where *j* = 0 or 1. We construct a 2×2 contingency table of counts of successes and failures as given by Collet (1991).

Table 3.3 A 2×2 contingency table with observed and expected values

	Group 0	Group 1	Total
Number of successes	$\boldsymbol{y_0}$	$\boldsymbol{y_1}$	$\boldsymbol{y_0 + y_1}$
Number of failures	$\boldsymbol{n_0 - y_0}$	$\boldsymbol{n_1 - y_1}$	$\boldsymbol{(n_0 + n_1) - (y_0 + y_1)}$
Toal	$\boldsymbol{n_0}$	$\boldsymbol{n_1}$	$\boldsymbol{n_0 + n_1}$

Adapted from: Collet (1991: page 34)

Let us denote an *observed* value in the i^{th} response for the j^{th} group as O_{ij} and a corresponding *expected* value as E_{ij}. Table 3.3 will then transform to a table giving observed *values* O_{10}, O_{11}, O_{20}, O_{21}, which correspond to expected values E_{10}, E_{11}, E_{20}, E_{21}. Expected values are calculated from the observations of Table 3.3 such as given in Table 3.4 below.

Table 3.4 2×2 contingency table showing calculation of expected values of successes and failures

	Group 0	Group 1
Successes	$\boldsymbol{E_{10}} = \dfrac{\boldsymbol{y_0(y_0+y_1)}}{\boldsymbol{n_0}}$	$\boldsymbol{E_{11}} = \dfrac{\boldsymbol{y_1(y_0+y_1)}}{\boldsymbol{n_1}}$
Failures	$\boldsymbol{E_{20}} = \dfrac{\boldsymbol{(n_0-y_0)\{(n_0+n_1)-(y_0+y_1)\}}}{\boldsymbol{n_0}}$	$\boldsymbol{E_{20}} = \dfrac{\boldsymbol{(n_0-y_0)\{(n_0+n_1)-(y_0+y_1)\}}}{\boldsymbol{n_0}}$

The χ^2 statistic is given by the formula

$$\chi^2 = \sum \frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

where O_{ij} is the observed value for the i^{th} level of the explanatory variable in the j^{th} category of the response variable, and E_{ij} is the expected value which is calculated as given in Table 3.3 (*i=0,1* and *j=1,2*).

Test of association:

The test of association is performed on ψ based on the assumption that the odds of success are the same for both groups. That is, the odds ratio (ψ) is approximately 1. We use the *5%* significance level for testing the null hypothesis against a composite (alternative) hypothesis that the odds for both groups are not the same (two-sided test). The value of the χ^2 is compared with a tabulated χ^2 value for 5% significance level on 1 *degree of freedom (Collet, 2003).* If the calculated χ^2 is greater than the tabulated value, the null hypothesis is rejected and we conclude that the odds of success for the two groups are not the same. Otherwise, do not reject H_0 and conclude that any difference in the odds of success between the two groups (short period of stay and long period) is due to chance alone.

Alternatively, we base rejection or non-rejection of H_0 on the probability value (*p-value*) of the 95% distribution of χ^2. If the *p-value* of the tabulated χ^2 is less than 0.05, H_0 is rejected in favour of H_1. It is, therefore, concluded that the *difference* in the odds of success between the two groups is *significant*; indicating that one of the groups is associated with one of the categories of the response variable.

3.1.5 The Logistic transformation

It is worth noting that in modelling, interest is centred on the distribution of p_i and Y_i not which is binary, i.e., is limited to only two values. Therefore, we will consider models for p_i, which can vary according to some values of an explanatory variable(s). We denote the explanatory variables $x_1, x_2, \cdots, x_k$. To emphasise that p_i changes with the x's, we write $p(x_1)$. A single explanatory variable is expressed as a random variable with *m* levels. If *m=2*, the model reduces to a 2×2 contingency table that was discussed in Sub-section 3.2.1.

The Logistic Regression model is the most suitable type of generalised linear regression models for analysing binomial response data. In this model, the dependence of the probability of success of the response data on explanatory variables is transformed from the range (*0,1*) to (-∞, ∞) (Collet, 2003, Hosmer and Lemeshow, 2000, Menard, 1995). Then a linear model is constructed for the transformed value of the success probability, ensuring that these values lie between 0 and 1. Ordinary linear models that are based on least squared method of estimation, do not consider this transformation and, therefore, yield misleading results, such as the values of the fitted probabilities lying outside the range 0 and 1! (Hosmer and Lemeshow, 2000).

The Logistic Regression model adopts one of three forms of transformation of the values of the success probabilities, namely (i) the *logistic* transformation, (ii) the *probit* transformation and (iii) the *complementary log-log* transformation. For a binary response variable, as according to Collet (2003), the logistic transformation is "from computational viewpoint, the logistic transformation is the most convenient". Two more other form of transformations are the *Negative Complementary Log-Log* and the *Cauchit* transformations which are suitable for the *Cumulative Link Models* (See SPSS, 2006)

The logistic transformation, also known as the logit, of success of probability p is $log\{p/(1-p)\}$ or simply *logit*(*p*). Recall from Section 3.2 that this is the log odds of a success. Note that any value of *p* in the range (*0, 1*) corresponds to a value of *logit*(*p*) in the range (-∞, ∞) – the domain for continuous and normally distributed data. Note also that, as *p→0* (that is, *p* tends to *0*), *logit*(*p*) → -∞ ; as *p→1*, *logit*(*p*) → ∞, and as *p=0.5*, *logit*(*p*) = 0

(McCullagh and Nelder, 1989). All these postulations lead to the fact that the function of *logit*(*p*) results in a sigmoid curve as displayed in Figure 3.2.

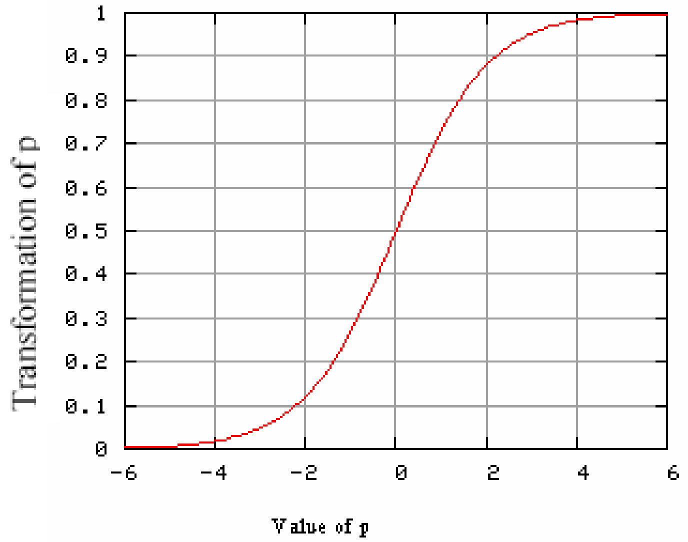

Figure 3.2: The logistic transformation of p (StatSoft, 2007)

3.2 The linear Logistic Regression model

As mentioned in Sub-section 3.2.1 above, the dependent variable in Logistic Regression is usually *dichotomous*, that is, the dependent variable can take the value 1 with a probability of success *p*, or the value 0 with probability of failure *1-p*. This type of variable is called a Bernoulli (or binary) variable. As will be explained in Chapter 4 the application of Logistic Regression is also extended to the case where the dependent variable is in form of ordered categorical responses This is also known as Ordinal Logistic Regression Model or Proportion Odds Model (McCullagh and Nelder, 1989).

The independent or predictor variables in Logistic Regression can take any form. That is, Logistic Regression makes no assumption about the distribution of the independent variables. They do not have to be normally distributed, linearly related or of equal variance within each group. The relationship between the predictor and response variable is not a linear function in Logistic Regression. Instead, the Logistic Regression function is used, which is the *logit transformation* of *p*. The derivation of the logit transformation rests on the fundamental theory of Generalised Linear Models (GLMs), which relate the probability of success of an outcome to the expected value of the outcome variable through a link function (McCullagh and Nelder, 1989) defined as

$$E(Y_i) = \eta_i(.),$$

where $\eta_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ik} + \sum_{i=1}^{k} \beta_j x_{ji};\ x_{0i} = 1; j = 0{,}1$ and $i = 1,2,..,n$.

A sample of size n and k explanatory variables is shown in Table 3.5, which in fact is an $n \times k$ contingency table. This illustration is meant to reinforce the understanding of the mathematical notation used in the subsequent explanation of the technical concepts.

Table 3.5 Layout of a sample of n observations, one response variable and k explanatory variables

Observation	**Response Variable**	**Explanatory Variables (X_{ji})**				**Residual Term**	**Probabilities of Success P(Y_i=1)**	
i	Y_i	X_{1i}	X_{2i}	$\cdots$	X_{ki}	ε_i	Observed*	Predicted
1	y_1	x_{11}	x_{21}	$\cdots$	x_{k1}	ε_1	$p_1 = \frac{e^{\eta_1}}{1+e^{\eta_1}}$	$\hat{p}_1 = \frac{e^{\hat{\eta}_1}}{1+e^{\hat{\eta}_1}}$
2	y_2	x_{12}	x_{22}	$\cdots$	x_{k2}	ε_2	$p_2 = \frac{e^{\eta_2}}{1+e^{\eta_2}}$	$\hat{p}_2 = \frac{e^{\hat{\eta}_2}}{1+e^{\hat{\eta}_2}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	y_n	x_{1n}	x_{2n}	$\cdots$	x_{kn}	ε_n	$p_n = \frac{e^{\eta_n}}{1+e^{\eta_n}}$	$\hat{p}_n = \frac{e^{\hat{\eta}_n}}{1+e^{\hat{\eta}_n}}$
Coefficients: Observed	β_0	β_1	β_2	$\cdots$	β_k			
Estimates	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\cdots$	$\hat{\beta}_k$			

Given that $E(Y_i) = n_i p_i$ and that p_i is the corresponding response probability, the linear logistic model for the dependence of p_i on the values of k explanatory variables, $x_{1i}, x_{2i}, \cdots, x_{ki}$ of $X_1, X_2, \cdots, X_k$ (Collet, 2003) is such that

$$Logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki}$$

This equation represents the definition of the Logistic Regression model. It means that in the logit scale, the probabilities of success are linearly related to the covariates $X_1, X_2, \cdots, X_k$. Algebraic rearrangement results in

$$p_i = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki})}{1 - \exp(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki})} = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad (3.9)$$

which is the form of equation in column 8 of the Table 3.4 representing the observed probability of observing a "success".

3.2.1 Fitting the linear Logistic Regression model to binomial data

In order to fit a linear logistic model to a set of data with *k* explanatory variables, we need to estimate the *k+1* unknown parameters $\beta_0, \beta_1, \cdots, \beta_k$. Since the form of the distribution of the data is binomial, we maximise the likelihood function of the distribution described in Subsection 3.1.1.

$$L(\beta) = \prod_{i=1}^{n} \binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i}$$

It is to be noted that this likelihood is a function of $\underline{\beta}$ (the vector of β's) since it depends on the unknown p_i that in turn depend on the β's through equation (3.9). Therefore, maximising the partial derivatives of the normal equations of the log likelihoods yields the maximum likelihood estimators of $\beta_0, \beta_1, \cdots, \beta_k$, *i.e,* $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$ (Hosmer & Lemeshow, 2000, p.33-35; Collet, 1991, p. 57). It follows that the relationship between the estimated response probability and the explanatory variables $x_1, x_2, \cdots, x_k$, can be expressed as

$$logit(p) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k,$$

which is rearranged to give

$$\hat{p}_i = \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki})}{1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki})}$$

Given that the estimated value of the linear systematic component of the model for the i^{th} observation is $\hat{\eta}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \cdots + \hat{\beta}_k x_{ki}$, which is also referred to as the linear predictor, the fitted probabilities can be found to be

$$\hat{p}_i = \frac{\exp(\hat{\eta}_i)}{1 + \exp(\hat{\eta}_i)}$$

which is the form of equation in column 9 of the Table 3.5 representing the predicted probability of observing a “success”.

These fitted individual probabilities are later compared with the observed probabilities for each group and each category of the response variable, to evaluate the validity of the fitted model.

As an illustration, we examine a statistical model that describes the relationship between the probability of a household becoming food insecure and sources of livelihood namely; agricultural production, livestock rearing, fishing, employment/labour, petty trade, food aid and other. Let us assume that interest centres on estimating the success probability of a household becoming food secure at each level of a source of livelihood. Fitting the data into a Logistic Regression model in SPSS will generate estimates of the *k+1* parameters described above as well as the estimate of the odds ratio and the predicted probability of “success” based on the ratio of the difference of two models: a model will all seven variables fitted and a model with only the response variable fitted. Hosmer and Lemeshow (2000) and Collet (2003) describe the steps for estimating of model parameters and testing an association between the k+1 explanatory variables and the response variable. Noting that all the variables are discrete or categorical each with two possible values, we use a collection of *design* or *dummy variables*. For each of the variables, we use the dummies, D_0 and D_1 for the values “No” and “Yes” respectively. If there were more than two levels for any of the explanatory variables, it would have k-1 design variables (Hosmer & Lemeshow, 2000, p.32; Collet, 1991, p.48).

3.2.2 The standard errors of parameter estimates

Standard errors of parameter estimates are needed to check the degree of precision of the parameter estimates. A standard error of an estimate of a Logistic Regression model denoted by $se(\hat{\beta})$, are readily given in SPSS and SAS outputs (SAS, 2008, SPSS, 2008). Consequently, we derive the *100(1-α)%* confidence limits for the corresponding true value of the individual parameter estimated. Thus the *100(1-α)%* confidence limits for β are $\hat{\beta} \pm \frac{z_\alpha}{2} se(\hat{\beta})$ where $z_\alpha/2$ is the upper *α/2* point of the standard normal distribution. SPSS and SAS also give the value of the *t*-statistic = $\hat{\beta}/se(\hat{\beta})$, which is used to test the hypothesis that the corresponding coefficient in the linear regression model is zero (Collet, 2003).

3.2.3 Testing for the significance of the model

Testing for significance of a model is the act of assessing the model to see how good it fits (other texts express it as *goodness of fit of a model*). It is best practice to investigate how the fitted values cmpare with the observed values, which act either require to be revised or accepted (Collet, 1991). In Logistic Regression the process usually involves testing for the significance of the *k* coefficients of explanatory variables (factors) using the *likelihood ratio test* based on the *statistic G* (Hosmer & Lemeshow, 2000) or *D statistic*, otherwise known as the *Deviance* (Collet, 1991, p.63).

The Logistic Regression model aims at testing the hypothesis that there is no difference between the levels of the prognostic factor with regard to the outcome variable, i.e. there is no advantage of one group over the other. The SPSS LOGISTIC REGRESSION or SAS PROC LOGISTIC procedure gives values of three tests, namely; the Likelihood Ratio Test, the Score Test and the Wald's Chi-square.

The Likelihood Ratio Test

The *likelihood ratio test* (LRT) is a statistical test of the *goodness-of-fit* between two models. A relatively more complex model is compared to a simpler model to see if it fits a particular dataset significantly better. If we establish that there is good fit, the additional parameters of the more complex model are often used in subsequent analyses. This test is only valid if used to compare *hierarchically nested models* (Collet,1991: p.68) That is, the more complex model must differ from the simple model only by the addition of one or more parameters. Adding additional parameters will always result in a higher likelihood score. However, there comes a point when adding additional parameters is no longer justified in terms of significant improvement in fit of a model to a particular dataset (Cox and Snell, 1989). The LRT provides one objective criterion for selecting among possible models. The LRT begins with a comparison of the likelihood scores of the two models. In other words, the LRT compares the *deviances* values of two models. The LRT is based on the change in the deviance of the model with the factor fitted and that of the model with only the intercept fitted. Hence, if the deviance for the model with the prognostic factor fitted is $D(\hat{\theta})$ and the deviance for the model with only the intercept fitted, i.e. $\hat{\theta} = \hat{\beta} = 0$, is *D(0).* The likelihood ratio (LR) is expressed as

$$LR = D(\hat{\theta}) - D(0) = -2\{\log L(0) - \log L(\theta)\}$$

where $-2\log L = -2\mathrm{l}(\hat{\theta})$ is the deviance of the model with factor from that without the factor i.e. the *null model* (Collet, 2003).

A description of the deviance follows in Section 3.3.4. The Likelihood Ratio statistic approximately follows a *chi-square* distribution (Collet, 2003, Hosmer and Lemeshow, 2000, McCullagh and Nelder, 1989). To determine if the difference in likelihood scores between the two models is statistically significant, we must also consider the degrees *of freedom*. In the Likelihood Ratio Test, degrees of freedom are equal to the number of the additional parameter in the model with a factor. Obviously there is only 1 degree of freedom since there is only one parameter in the model with one factor. Using this information we can then determine the critical value of the test statistic from standard statistical tables.

The Score Test

The score test is the equivalent of the Pearson's chi-square statistic described in Chapter 3. It is defined as

$$\chi^2 = \frac{Z^2}{V}$$

The Wald Test

The Wald test is defined as

$$\chi^2 = \left[\frac{parameter\ estimate}{se(estimate)}\right]^2$$

3.3 The Logistic Regression Model for Ordered Categorical Data

In Section 3.2 methods for exploring relationships between independent variables and a dichotomous dependent variable were presented. In this Section, the response variable has three or more levels or categories which are ordered from a lowest-ranked category to the highest. In the case of response variable *food consumption score,* there are three categories namely; *Poor, Borderline poor* and *Good.* There are a number of methods available to model ordered categorical data – *linear-by-linear models, continuation ratio logits* and *proportional odds* are some of the more widely used (Collet, 2003, Hosmer and Lemeshow, 2000).

However, because of the nature of the data available for this project, only the proportional odds model is discussed here.

The proportional odds model can be understood as an extension of Logistic Regression or, as proposed by Collet (2003), a "generalisation of the Logistic Regression". Therefore, calculation and interpretation of the model parameters and *deviance* statistics is the same as for the Logistic Regression for binary response data. The technique allows one to model ordered data by converting the data into a number of dichotomies. A binary Logistic Regression models one dichotomy whereas the proportional odds model uses a number of dichotomies. The ordered data are arranged as a series of binary comparisons. For the dataset of this project, a three-category ordered variable (coded 1, 2, 3) is represented as two comparisons: (a) Category 1 compared to categories 2 and 3; (b) Categories 1 and 2 compared to category 3. Such method of modelling is referred to as all possible Logistic Regression model (see Collet, 2003: p. 325-6).

3.3.1 Formulation of the proportional odds model for ordered categorical data

The proportional odds model, abbreviated POM, of a relationship between *m* independent variables each with *h* levels and one response variable with k ordered categories is derived by Collet (2003: p. 325-9). In this text, the *k* ordered categories of the response variable Y, are denoted by $C_1, C_2, \ldots, C_k$, where $k \geq 2$ and where a response in category C_j can be described as "worse than" one in $C_{j'}$, if $j < j'$. Thus for the food consumption score with the responses (or outcomes) labelled as "poor", "borderline poor" or "good", the categories would be C_1, C_2 and C_3 so that $C_1 < C_2 < C_3$.

If we suppose that Y_i is a categorical response variable for the i^{th} household with *k* levels, it turns out that Y_i takes the value *j* if the response is in category C_j, $j=1,2\ldots,k$. If x_i denotes a value of an explanatory (or independent) variable X, the probability that the i^{th} household responds in category C_j, is denoted by p_{ij}, such that

$p_{ij} = P(Y_i = j)$ = P[household *i* responds in category *Cj*].

It follows that, the cumulative probability of a response in category *Cj* or worse, denoted as $\gamma_{ij}$is $\gamma_{ij} = p_{i1} + p_{i2} + \cdots + p_{ij}$. As a result, $\sum_{j=1}^{k} p_{ij} = \gamma_{ik} = 1$ (Collet, 2003).

More understanding of the theory on formulation of the Proportional Odds Model (other texts use 'Cumulative Odds Ratios ') can be found in Agresti (2002), McCullagh (1980), Peterson and Harrel (1990) and SPSS (2006).

3.3.2 Comparison between two households

One of the advantages of fitting a multiple Logistic Regression model is that it enables us compare between households. Suppose we have two households *a* and *b*, we denoted by h_a and h_b respectively. Comparison yields:

For household a: $\log\left\{\frac{Q_j(\underline{z}_a)}{1-Q_j(\underline{z}_a)}\right\} = \alpha_j + \eta(\underline{z}_a)$ and for household b: $\log\left\{\frac{Q_j(\underline{z}_b)}{1-Q_j(\underline{z}_b)}\right\} = \alpha_j + \eta(\underline{z}_b)$. The log-odds ratio of h_a relative to h_b is estimated as

$$\hat{\theta} = \log\left\{\frac{Q_j(\underline{z}_a)}{1-Q_j(\underline{z}_a)}\right\} - \log\left\{\frac{Q_j(\underline{z}_b)}{1-Q_j(\underline{z}_b)}\right\} = \log\left\{\frac{Q_j(\underline{z}_a)}{1-Q_j(\underline{z}_a)}\right\}\Big/\log\left\{\frac{Q_j(\underline{z}_b)}{1-Q_j(\underline{z}_b)}\right\}$$

$$\log\left\{\frac{Q_j(\underline{z}_b)}{1-Q_j(\underline{z}_b)}\right\} = \log\left[\frac{Q_j(\underline{z}_a)\{1-Q_j(\underline{z}_b)\}}{Q_j(\underline{z}_b)\{1-Q_j(\underline{z}_a)\}}\right] = \left[\alpha_j + \eta(\underline{z}_a)\right] - \left[\alpha_j + \eta(\underline{z}_b)\right]$$
$$= \eta(\underline{z}_b) - \eta(\underline{z}_b)$$

3.3.3 The Mann-Whitney test of the proportional odds model

The Mann-Whitney test statistic used in this text is based on the efficient score (*Z*) for the common odds ratio, θ, and Fisher's information *(V)* as derived by Jones and Whitehead (1979). The *efficient score θ* is given as

$Z = \frac{1}{n+1}\sum_{j=1}^{k} a_j(B_{j-1} - B^{j+1})$

Where a_j is the number of individuals in group *0* observed in category *j*, B_{j-1} is the number of individuals in previous categories of group *1*, B^{j+1} is number of individuals in subsequent categories of group *1*.

Fisher's information is given by $V = \frac{n_0 n_1 n}{3(n+1)^2}\left\{1 - \sum_{j=1}^{k}\left(\frac{t_j}{n}\right)^3\right\}$

where t_j is total number of households in category *j*.

For large samples and small θ, $Z \sim N(\theta V, V)$, i.e. *Z* is approximately normally distributed with mean θV and variance *V*. Under H_0: $\theta=0$, Z^2/V is referred to a χ^2 distribution under 1 degree of freedom. The log odds ratio, θ, is therefore estimated by $\hat{\theta} = Z/V$, which is an *approximate* maximum likelihood estimate with an approximate 95% confidence interval given by

$$\left(\frac{Z}{V} \pm 1.96\frac{1}{\sqrt{V}}\right).$$

3.3.4 Fitted Probabilities

From theory, the log-odds of *normal* to *moderate or worse* is given by

$$\log\left\{\frac{Q_j(\underline{z}_a)}{1 - Q_j(\underline{z}_a)}\right\} = \alpha_j + \eta(\underline{z}_a)$$

Or

$$\frac{Q_j(\underline{z}_i)}{1 - Q_j(\underline{z}_i)} = e^{((\alpha + \eta(\underline{z}_i))}$$

And finally,

$$Q_j(\underline{z}_i) = \frac{1}{1 + e^{-(\alpha_j + \eta(\underline{z}_i))}}$$

3.3.5 Calculating the Deviance

Selection of important variables to include in a *k*-category ordered model depend on the values of the deviance of models. According to Ashby, Pocock and Shaper (1986), “for two models with the same subjects, one with *p* independent variables, and the other with an extra *q* independent variables, twice the difference in the maximised log likelihood is distributed asymptotically as χ^2 on q degrees of freedom, under the null hypothesis that the extra q variables do not discriminate between variables”. The raw deviance calculated from grouped binary data, is defined as

$$D = -2\log L = -2\sum_{ij} w_{ij} \log \hat{p}_{ij}$$

Where $\widehat{p}_{ij}$= fitted probability for j^{th} cell in group *i* and $\widehat{w}_{ij}$= weight (count) for j^{th} cell in group *i*. The *deviance* is used in calculating the likelihood ratio statistic for testing the null hypothesis of no difference between groups.

3.3.6 Hypothesis testing

To test $H_0: \theta = 0$, where θ is the common value of the odds ratio, we consider the likelihood ratio test statistic as given by the difference of deviances (*see mathematical definition of deviance in subsection 3.3.5*):

$$\chi^2 = D(0) - D(\theta) = -2\ell(0) - [-2\ell(\hat{\theta})]$$

SPSS and SAS output the value of the *Deviance*.

3.3.7 Model Checking

After selecting a model, it is important to check whether proportional odds model is appropriate for the data. The techniques that follow involve examination or diagnosis of fitness of the proportional odds (parallel regression lines for *cumulative logits*) assumption by looking at the case of grouped data. Fitted probabilities and testing the proportional odds assumption is looked at in the following sub-sections. The text on model checking will be mainly based on Collet (2003: Chapter 5) who methodically discusses how the Logistic Regression assumptions are inspected.

Score test for the proportional odds assumption

The SAS output includes a result of a special test featuring the Score Test. A significant chi-square value indicates lack of fit of the proportional odds assumption, while a non-significant test shows goodness of fit and the hypothesis that the regression lines for cumulative *logits* are parallel cannot be rejected. A significant chi-square test indicates that the proportional odds assumption is not justified.

Fitted probabilities and frequencies

As we now know that the proportional odds model uses cumulative *logits* of ordered categorical data, there is need to inspect how close are the fitted cumulative proportions of each of the three categories to the observed proportions (see Agresti, 2004, Ashby et al., 1986, Collet, 2003). Apparently, close fitted and observed proportions indicate the fitted model is good for estimating relationships. Like Logistic Regression, ordered *logit* uses maximum likelihood methods, and finds the best set of regression coefficients to predict values of the *logit*-transformed probability that the dependent variable falls into one category rather than another. Logistic Regression assumes that if the fitted probability, *p*, is greater than *0.5*, the dependent variable should have value *1* rather than *0*. Ordered *logit* doesn't have such a fixed assumption. Instead, it fits a set of cut-off points. Because there are 3 levels of the dependent variable visual acuity (*1* to *3*), it will find *3-1 = 2* cut-off values k_1 to k_2 such that if the fitted value of *logit(p)* is below k_1, the dependent variable is predicted to take value *0*, if the fitted value of *logit(p)* is between k_1 and k_2, the dependent variable is predicted to take value *1*, and so on. As with Logistic Regression, we get an overall chi-square for the goodness of fit of the entire fitted model, and we can also use a chi-squared test to assess the improvement due to adding an extra independent variable or group of independent variables as with Logistic Regression, a crucial piece of information for evaluating the fit of the model is a table of predicted versus observed category.

CHAPTER FOUR

DATASET AND METHODOLOGY

4.0 Introduction

In this Chapter the dataset (or the sample) and the methodology employed in data analysis, which is done in the next Chapter, is described. The first step is to describe the dataset used in the study. Secondly, the methodology used in the survey which produced the dataset, i.e. sample selection and survey field work protocols, is described. Third, the methods of deriving the data for the response variable Food Consumption Score (FCS) are described. Fourth, the steps leading to analysis of data, i.e. Proportional Odds Model (POM) of the Logistics Regression techniques family for selecting the variables of importance, are explained. Lastly, we present the methods used in *Model Diagnostics*. The main aim of this Chapter is to stimulate an appreciation of the relevance and robustness of the statistical techniques used in the analysis (the subject of Chapter 5).

4.1 Description of the dataset

As mentioned above, the raw dataset used for the study was obtained from the Sudan Household Health Survey (2006). The survey was modelled on Multiple Indicator Cluster Survey (MICS) methodologies. The United Nations Children's Fund (UNICEF) adopts MICS procedures for carrying out surveys for measuring progress towards the Millennium Development Goals (MDGs) and in particular, the situation of women and children in a specific country (UNICEF, 2008). MICS uses a set of modular questionnaires to cover the broad sectors of health and education. In the case of the 2006 Sudan Household Health Survey (SHHS), food security was also covered. In total, five questionnaires were administered, namely: Community, Household, Women, Under-five and Food Security Questionnaires. The Food Security Questionnaire (see Appendix I) covered seven sections or modules, namely: household circumstances; household belongings and livestock; livelihoods and agricultural production; household expenditures; food consumption and sources; shocks and coping mechanisms and food aid. The survey aimed to derive key indicators from the seven information categories: (i) household characteristics; (ii) nutrition; (iii) child health; (iv) water and sanitation; reproductive health; (v) HIV/AIDS; (vi) education and; (vii) food security.

4.2 Sample selection, data collection and processing

Sudan is administratively divided into 25 federal states. Ten of the states are in Southern Sudan, which acquired an autonomous governance status following the conclusion of a peace agreement on 9th January 2005. Each state comprises a number of counties. The SHHS used states as *domains* of data analysis. Sampling was based on the multi-stage stratified sampling design. One thousand households were selected per state, making a total of 25,000 households for the whole country. Each state was divided into segments or villages from which 40 were randomly selected. Then 25 households per segment/village were selected at random. A detailed description of the sampling methodology is described in the survey report (GoSS-MOH, 2008: p.235-245). Coverage by enumeration resulted in 24,527 respondents or 98% response rate. A total of 9,557 households were enumerated in Southern Sudan representing a response rate of 95.6%. However, when examining the raw dataset 337 forms (3.5%) were entered with only the identification information and the field "Result of Interview" entered as "Not at home", "Refused", "Household not found/destroyed" and "Other". The values with these entries, amounting were removed and the working sample size is now 9,220 households.

Given the political set up of Sudan (as prescribed in the Comprehensive Peace Agreement of 2005), training and field work were conducted separately in Khartoum, the nation's Capital, and the town of Rumbek in Southern Sudan. Field work took almost three months from March to May 2006.

During the Survey, a team of interviewers led by a supervisor visited selected villages, parts of villages or segments (in major towns) that were included in this survey. Enumeration (interviewing) of households was preceded by a household listing operation. In this operation, interviewers listed all households in the sample segment, marking a number on each household in chalk and asking several questions at each household according to the Village Listing Sheet. The household listing exercise is important to determine the number of households to be covered in the survey and number of households in certain house blocks or big buildings. The list was also used to facilitate supervisor control functions. Then interviewers returned to selected households, as indicated by the supervisor, to administer four of the questionnaires except the Community Questionnaire.

For the Food Security Questionnaire, which is the subject of this study, the survey team was to interview a mother or a principal caretaker of children under-five years (in case of orphaned children). Where no eligible respondent was met, only the Household Questionnaire was completed by asking any adult present in the household to respond to the questions. In the event that the household appeared to have no one, a neighbour was asked if it was inhabited. If the household was occupied, the survey team would ask the neighbour when the household members would return. Arrangement would then be made with the supervisor to return to the dwelling at a later time or at the end of the day when its members returned. In case the household remained without any member on the subsequent visit, the enumerator would make a “Not at home” mark on the questionnaire. No household was to be substituted in the sample.

After all field operations were completed, data processing was conducted starting with training of 22 data processing personnel. The data processing package for population surveys, CSPro, was used for entering, controlling and tabulating data. Then the raw dataset was migrated into SPSS for facilitation of analysis.

4.3 Derivation of the main response variable

The main response variable under study is the Food Consumption Score (FCS). The Food Consumption Score is based on the dietary diversity, food frequency and relative nutritional importance (WFP-VAM, 2008) is a measure (indicator) of dietary diversity of food consumed in a defined period by a household or geographical area. It indicates the availability and consumption of specific food groups to determine the extent of nutritional vulnerabilities and vis-à-vis the level of food insecurity in an area. According to Swindale and Bilinsky (2006), the Household Dietary Diversity Score measures food access as a proxy of socioeconomic status. Twelve food groups are identified: (1) cereals; (2) roots and tubers; (3) vegetables; (4) fruits; (5) meat, poultry, offal; (6) eggs; (7) fish and sea food; (8) pulses/legumes/nuts; (9) milk and milk products; (10) oil/fats; (11) sugar/honey and (12) miscellaneous.

The Food Security Questionnaire of the Sudan Household Health Survey (see Appendix I) included eight questions on food consumed. This section asked (a) the number of times in a day household members (children and adults separated) consumed food during normal periods and in the hunger period, (b) number of times certain types of foods were consumed in the past week, (c) whether more, same or less amount of food was consumed when compared to the last harvest, (d) whether more, same or less amount of food was consumed when compared to the last rainy season and (e) the main source of the food item (i.e. own production, market purchase, surrounding natural resource, labour wage, borrowing, food aid, gift and others).

A number of steps lead to the calculation of the variable of the Food Consumption Score or Household Dietary Diversity Score. The questionnaire lists the food groups and food items as outlined in Table 4.1 below.

Table 4.1 List of food items per food group

	Food Group	**Food Item and examples**
1	Cereals and tubers	Maize, sorghum, millet, rice, other
		Cassava, potatoes, yam, other
2	Legumes or pulses	Beans, soya, groundnuts, tree nuts, seeds, other
3	Fruits and vegetables	Vitamin A rich vegetables (green leafy vegetables, yellow sweet potato, carrot, pumpkin, other)
		Other vegetables (tomato, cucumber, onion, other)
		Vitamin A rich fruits (mango, papaya, other)
		Other fruits (banana, apple, pineapple, other)
4	Animal protein products	Meat (beef, pork, lamb, game)
		Fish
		Poultry
		Other (eggs, rodents, insects)
5	Dairy products	Milk, yoghurt, cheese, cream
6	Oils and fats	Fats, oil, butter
7	Other	Sugar, salt, tea and other beverages

The FCS is calculated for each household based on the standards set out in Table 4.2. The number of food groups consumed by members of the same household is aggregated. The number of times a food item is eaten in a week or the frequency of food consumption and the standard weight of the food group provide the basis for calculation of the FCS.

Table 4.2 Standard food groups and standard weights for calculation of the Food Consumption Score

	Food Consumption Group	Food Group	Weight (definitive)
1	Maize , maize porridge, rice, sorghum, millet pasta, bread and other cereals	Main staples	2
	Cassava, potatoes and sweet potatoes, other tubers, plantains		
2	Beans. Peas, groundnuts and cashew nuts	Pulses	3
3	Vegetables, leaves	Vegetables	1
4	Fruits	Fruit	1
5	Beef, goat, poultry, pork, eggs and fish	Meat and fish	4
5	Milk yogurt and other diary	Milk	4
6	Sugar and sugar products, honey	Sugar	0.5
7	Oils, fats and butter	Oil	0.5
8	Spices, tea, coffee, salt, fish power, small amounts of milk for tea.	Condiments	0

Source: WFP-VAM (2008)

Using the SPSS® script editing module called '*Syntax Editor*', food consumption groups (FCGs) were obtained by aggregating individual frequencies of food items in the same group. The Food Consumption Score was then calculated for each household by summing up the product of the frequency of the FCG multiplied by the corresponding weight. The next step was to determine the thresholds for the FCS based on the frequency of the scores and the following criteria:

Table 4.3 Profiling of food consumption behaviour based on the Food Consumption Score

Food Consumption Score	Food Consumption Profile
≤ 28	Poor
28.1 - 42	Borderline
42.1 – 105	Good

Finally, the food insecurity level is concluded for each household based on the food consumption profile such that "*Poor*", "*Borderline*" and "*Good*" food consumption behaviour could mean poor, borderline and acceptable food security levels respectively.

4.4 The set of predictor variables

The food security dataset of the Sudan Household Health Survey includes a number of explanatory variables that are assumed to be associated with the food security outcome variable FCS. The set of (factors, independent or explanatory variables) to be investigated is included in Appendix 2. Nineteen predictors will be thrown into a Logistic Regression model and the reasons for doing so are stated. The selected set of predictors include three factors with quantitative (or ratio scale) values, otherwise referred to as *covariates* in this text. A number of other possible covariates have been purposefully left out of the model because of the numerous recall and enumeration errors or many missing values. Enumerators normally fail to quantify the measured items while respondents, who are mostly illiterate (See Chapter 5), are riddled with recall-bias. The variables not included in the model should not be regarded as not being useful in the study. They help in the interpretation of the data. Some of the variables included were for the uses of food aid and relief interventions, such as the UN World Food Programme and the FAO, to help in their planning tasks. Yet, other variables were excluded in order for the analysis not to violate the assumptions of *multicolinearity*, *non-additivity* and *heteroscedasticity* (see Hosmer and Lemeshow, 2000, Menard, 2002)

4.5 The data analysis techniques

The project work is to motivate further use of statistical modelling techniques in similar food security surveys as the Sudan Household Health Survey or Demographic and Health Surveys. The package SPSS will be used in determining the variables of interest, such as in the calculation of the Food Consumption Score, exploratory analysis, finding the important exploratory (or predictor) variables and production of model parameter estimates.

The SPSS technique PLUM will be used to: (i) investigate the goodness of fit of the model; (ii) generate parameter estimates for determining difference between categories of the response variable; (iv) calculate fitted probability values; (v) give model inspection goodness of fit statistics and; (vi) produce tests of hypothesis of the significance of the relationships or association between categories of the response variable and levels of the significant predictors. Tests of hypothesis will be based on (a) the Likelihood Ratio Test, (b) the Score Test and (c) the Wald's Chi-Square Test as described in Section 3.2.6.

SPSS analysis will involve the application of the Logistic Regression technique called Ordinal Regression, which is appropriate for the Proportional Odds Model. Interpretation of the outputs of the Ordinal Regression procedures will enable an in-depth understanding of the results and findings.

4.6 Model selection

The aim of this section is to investigate a model and possible predictors adjusting for one another and to find out whether any one predictor would still remain significant or non-significant when added to a model that already has one or more predictors. This procedure invites a sequence of likelihood ratio tests that help in observing whether any model with more than one predictor added to another model that already has other variables, would reduce the value of the deviance significantly. Besides testing for significance of the relationship between levels of fitted explanatory variables and the response variable, it is advisable to start looking at the magnitude of any relationship in terms of the odds ratio estimates.

Method selection allows specification of how independent variables are entered into the analysis (SPSS, 2006). Three methods are often used by researchers in investigation of relationships using logistic regression modelling. These are: forward selection, backward elimination and stepwise selection. SPSS version 16 includes under its LOGISTIC REGRESSION and REGRESSION procedures an option called "Method" that has seven options: Enter; Forward: Conditional; Forward: LR; Forward: Wald; Backward: Conditional; Backward: LR and; Backward: Wald for all three methods. These seven methods are summarised into three namely: Forward Selection; Backward Elimination and Stepwise Selection. It may be useful to discuss the advantages and steps of applying each of the methods and then explain why one of them was selected. It should be noted that the Proportional Odds Model or the SPSS PLUM or Ordinal Regression procedure uses the Stepwise method for determining significant and non-significant variables. For binary and continuous scale data the SPSS procedures mentioned above allow specification of the option “Method” in their option dialogues.

4.6.1 Forward Selection

The forward-selection technique begins with a model with no variables, also known as the *null model*. For each of the independent variables, the method calculates *F*-statistics that reflect the variable's contribution to the model if it is included. The *p*-values for these *F*-statistics are compared to the *SLENTRY* = value that is specified in the *model* statement (or to 0.50 if the *SLENTRY* = option is omitted). If no *F*-statistic has a significance level greater than the *SLENTRY*= value, the operation stops. Otherwise, the method adds the variable that has the largest *F*-statistic to the model. It then calculates *F*-statistics again for the variables still remaining outside the model, and the evaluation process is repeated. Variables are added one after the other to the model until no remaining variable produces a significant *F*-statistic. Variables that remain in the model are considered important predictors of the response variable.

4.6.2 Backward Elimination

The backward elimination method begins by fitting a full model, i.e. including all of the explanatory (independent or predictor) variables in the model and calculating *F*-statistics. Then the variables are dropped from the model one by one until all the variables remaining in the model produce *F*-statistics significant at the 0.10 level. The backward elimination method will stop when all the variables remaining in the model produce *F*-statistics with *p*-values less than the cut-off.

4.6.3 Stepwise Selection

In stepwise selection variables are added as in forward selection, but after a variable is added, all the variables in the model are candidates for removal.

Of the three methods of model selection, the backward elimination method is chosen simply because of automation. The software used (SPSS and SAS) automatically and effortlessly drops those variables which yield an F-Statistic smaller than the cut-off value. Further advantages of the backward elimination method over the other criteria, are described in the linear regression literature. According to Hocking (2003), both the forward selection and the backward elimination criteria have lent themselves to criticism. He intimates that the forward selection criterion is seen to be weak in that once a variable is entered it cannot be removed. Similarly, for the backward elimination algorithm once a variable has been removed it cannot be included. Singh (2004) notes that the limitation of both methods is in that once a variable is removed, their significance might change. The model would have thus been denied the inclusion of the removed variables.

4.7 Procedures for model checking and diagnostics

An important requirement in analysis using Logistic or Linear Regression modelling is that the model selected must be checked or diagnosed. That is, the model must be assesses to determine whether proportional odds is appropriate for modelling the data. In other words, we must examine the fit of the Proportional Odds Model (POM) assumption with the requirement that the plotted cumulative *logits* must be parallel. The process involves inspection of fitted probabilities and testing of hypothesis surrounding the proportional odds assumption.

4.7.1 The Score Test for validation of the proportional odds assumption

The Score Test features in the SPSS output. A significant Chi-Square value indicates lack of fit of the proportional odds assumption, while a non-significant test shows goodness of fit and the hypothesis that the regression lines for cumulative *logits* are parallel cannot be rejected.

4.7.2 Fitted probabilities and frequencies

As it is now known that the proportional odds model uses cumulative *logits* of ordered categorical data, there is need to inspect how close are the fitted cumulative proportions of each of the three categories to the observed proportions. Apparently, close fitted and observed proportions indicate the fitted model is good for estimating relationships. Like the simple logistic regression for binary response data, ordered *logit* uses maximum likelihood methods, and finds the best set of regression coefficients to predict values of the *logit*-transformed probability, that the dependent variable falls into one category rather than another. Logistic regression for binary data assumes that if the fitted probability, *p*, is greater than 0.5, the dependent variable should have the value 1 rather than 0. Ordered logit doesn't have such a fixed assumption. Instead, it fits a set of cut-off points. Coming to our dataset, because there are 3 levels of the dependent variable FCS (1 to 3), it will find 3-1 = 2 cut-off values k_1 to k_2 such that if the fitted value of $logit\ (p)$ is below k_1, the dependent variable is predicted to take value 0, if the fitted value of $logit\ (p)$ is between k_1 and k_2, the dependent variable is predicted to take value 1, and so on. As with logistic regression for binary data, we get an overall Chi-Square for the goodness of fit of the entire fitted model, and we can also use a Chi-Square test to assess the improvement due to adding an extra independent variable or group of independent variables. A crucial piece of information for evaluating the fit of the model is a table of predicted versus observed category. All these calculations require the use of SPSS and will be shown in the next Chapter.

Recalling that there are three categories of FCS, a category of FCS is denoted as C_j , where *j* = 1, 2 or 3. That is, C_1, C_2 and C_3 for *poor*, *borderline poor* and *good* food consumption. SPSS enables comparison of the fitted probabilities with the observed probabilities using a procedure known as Classification Table. The predicted probabilities are included in a variable added to the dataset which SPSS creates This variable is then cross-tabulated with the actual probabilities to determine the percentage of correctly predicted categories of food consumption. It can be concluded that the more the amount of correctly predicted categories, the better the goodness of fit of the model. An alternative approach is to examine the *observed* and *expected* frequencies. We multiply the fitted probabilities by the frequencies of the respective number of households in that category of FCS to obtain the expected frequencies. In interpreting the findings, if the *observed* and *fitted* probabilities are close for the predictor variables examined, it reflects goodness of fit of the model. Otherwise, if this is not the case, it could be that the distribution of the observations between the categories of the predictor variables compared is disproportionate.

4.7.3 Direct assessment of the model assumption for the proportional odds model

The proportional odds model assumption is that $\underline{\beta}_1 = \underline{\beta}_2 = \cdots = \underline{\beta}_{1-k}$ i.e., the parameters corresponding to each predictor are the same for each dichotomisation of the data. To test this assumption, it is necessary to set up (*k-1*) binary data sets. Hence, since we have three ordered categories of the response variable, we set up 2 data sets as follows: (1) Dataset 1 is constructed by putting category 1 as *success* and category 2 and 3 as *failure*. (2) Dataset 2 is obtained by putting categories 1 and 2 as success and category 3 as failure. Then the same linear model will be fitted to each of the two data sets and the $\underline{\beta}$ estimates are compared.

CHAPTER FIVE

DATA ANALYSIS, RESULTS AND DISCUSSION

5.0 Introduction

This chapter is concerned with analysis of data applying Logistic Regression techniques on the 2006 Sudan Household Health Survey dataset that has Food Consumption Score (FCS) as the variable of interest. Food consumption score is a proxy measure of food in/security in a given geographical location (Singh, 2004, WFP-CFSVA, 2007, WFP-VAM, 2008). The study end is to explore and determine factors and covariates that predict the outcome of food consumption score. In less technical jargon, analysis will focus on establishing presence or absence of relationships between a set of selected factors and food consumption score. In this text, *FCS* will be the dependent (Y) and referred to as the *response variable*, while a set of selected independent variables (*factors* and *covariates*) will be referred to as *explanatory variables* (X).

The Chapter is divided into three sections with each section subdivided into sub-sections. Section 5.1 will cover exploratory and descriptive statistical methods. There will be three approaches for exploring the data. Subsection (1) will examine the distribution of the response variable food consumption score (FCS) when it is a continuous variable. This set of analyses will be based on measures of central tendency and dispersion, frequency distributions, probability distribution plots and box plots. Subsection (2) will be on exploration of relationship involving FCS and each of three explanatory variables measured on a continuous scale (numeric), and finding correlation statistics (i.e. Pearson's R and Spearman's Correlation statistics for testing significance of two-way relationships). Subsection (3) will involve the use of simple $n \times m$ cross-tabulations for determining the significance of a 2-way relationship with Pearson's Chi-Square (χ^2) values, and probabilities. Pearson's Chi-Square indicates whether relationships exist between categories of the ordered response variable and levels of explanatory variables. Section 5.2 will use the logistic regression technique known as Proportional Odds Model (POM) (Agresti, 2004) or simply

Ordinal Regression (SPSS, 2006). As described in Chapter 4, the proportional odds model is used when the variable of interest (the response variable) has categories ranked according to a natural order. Recall that the grouped *FSC* values are in the form of three ordered categories: *poor, borderline* and *acceptable* (or *good*) consumption ranked as 1, 2 and 3 respectively.

It is also possible to analyse the data using the Ordinary Linear Regression (OLR) or the Multiple Linear Regression (MLR) models. In so doing, the calculated values of *FSC* will be treated as those of a continuous or interval scale variable. However, the interest of this project is to demonstrate the use of Logistic Regression that models the grouped values of food consumption scores (i.e. the variable *FCG* instead *of FCS*). The rationale behind this choice is that interest is centred on comparison of the three ordered categories of food consumption scores. Nevertheless, the reader might be interested in knowing how the results of the Logistic Regression techniques compare with those of the Linear Regression model. For this reason, annotated analysis using the Linear Regression technique will be presented and compared (Section 5.3). As explained in Chapter 4, the backward elimination strategy will be used for selecting the predictor variables of influence. Variables that are seen not to be contributing significantly to the model will be dropped and the model fitted again. The predictor variables selected are both factors (categorical) and covariates (interval scale).

Analysis will be based entirely on the statistical software package SPSS®. Modelling will be based on two procedures called *PLUM* and *REGRESSION* for the Ordinal Logistic Regression and Linear Regression, respectively. Results from the electronic outputs of the procedures will be discussed and appropriate interpretations of useful findings made.

Nineteen predictor (independent or explanatory) variables will be examined and included in the model, although the survey from which the dataset was derived includes many possible predictor variables. However, this set of variables is not small by any standard. The rest of the possible predictors were dropped on three grounds. In the first case, they contained numerous missing values, which is a result of two possible causes: (i) Lapses in recall of measurements and quantities of such assets as area under cultivation, quantity of harvested crops, quantity of food consumed or amount of income earned or spent; (ii) skipped questions as a result of the nonapplicable responses. For example, if the response to the question: “Does your household usually use land for farming” include a big number of respondents saying ‘No’, such as in the case of non-farming pastoralist communities, internally displaced people and urban dwellers, the response to the next question asking about amount of land under cultivation, automatically does not apply. The second case of missing values arises from recall errors. This is not to be unexpected since the sample size included a big number of respondents with no or low literacy (45.8%). For this category of respondents, record keeping is not known or practiced. Even if enumerators were trained on how to elicit the answers by probing and imputing, it would be impossible in most cases or time consuming.

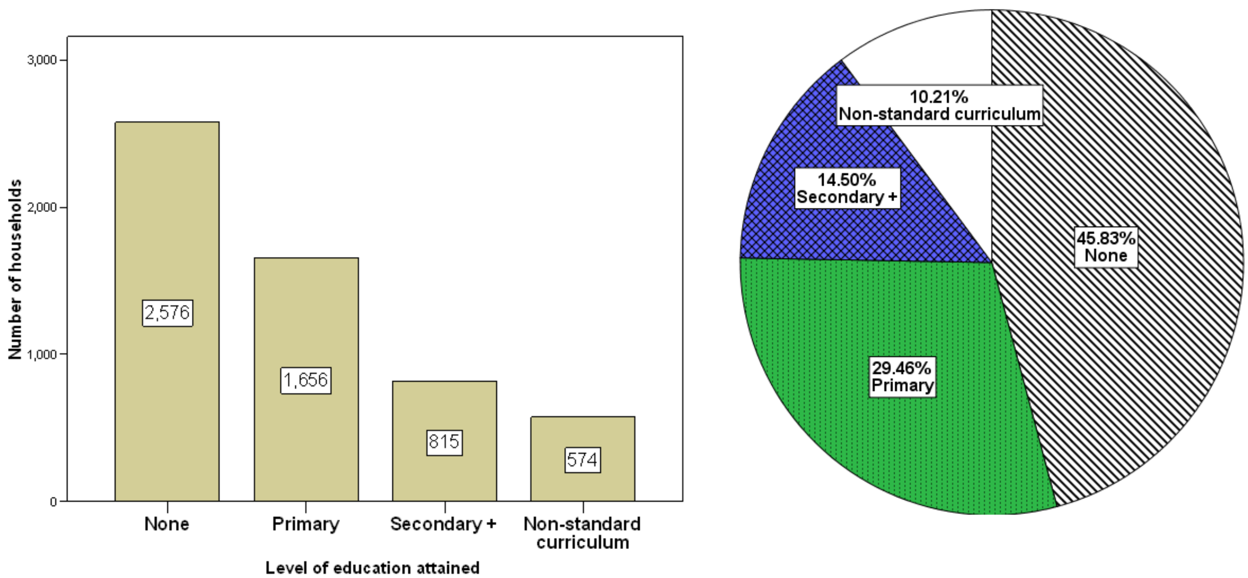

Figure 5.1: Bar and pie charts showing number and percentage of households by level of education attained

As Figure 5.1 above shows, only 14.5 *per cent* of household heads interviewed completed secondary school or higher level of education.

5.1 Exploratory analysis

Exploratory analysis uses descriptive statistical methods for giving the first impression about the data and relationships between the dependent and independent variables. This section features selection of predictors and presents three approaches to the exploratory analysis: (a) exploring the distribution of the data based on the response variable with continuous values; (b) exploring relationship between two numerical predictors by determining their correlation coefficients and their significant probabilities and; (c) exploring possible relationships between the response variable *FCS* and the set of 19 selected predictors. It is noteworthy that the set of selected predictors includes three *covariates* i.e. number of household members (household size), number of months in which harvested food lasted and number of meals eaten per day (by both adults and children during normal, non-hunger period). Approach (b) will only apply to these three *covariates*.

Before delving into methodical analysis of the data, it could be important to have a quick glance of the variables that are included in the study. Table 5.1 lists the variables by name, number of levels (categories), type, and the measure scale according to SPSS nomenclature.

An important variable derived by calculation, is the Wealth Index Scores (*WIS*) with values classified into quintiles (i.e. 20th percentiles of the wealth index) to create Wealth Index Quintiles (*WIQntile*). The Wealth Index Scores and *WIQntile* were extracted from the original survey dataset. The calculation of the index is based on ownership of assets and weights. The procedure is explained in the 2006 Sudan Household Health Survey report as well as in WFP Comprehensive Food Security and Vulnerability Analysis report (WFP-VAM, 2007a). Basically, households are classified according to five ordered categories: *poorest, poorer, moderate, richer,* and *richest* depending on how they rank on the quintile scale. With only 1.7 *per cent* missing values, it is found out that 71 *per cent* of the households are of *poorest* to *moderate* of wealth index bracket. This result does not come as a surprise given the post-conflict setting of Southern Sudan and how Sudan ranks in terms of the Human Development Index (HDI). The country ranks amongst the 50 least developed countries (UNDP, 2009). As wealth index is an indirect measure of food in/security, it is highly important to examine the statistics of the relationship between food consumption score and wealth index quintile. Hence, the ordinal variable wealth index quintile is included among the 19 selected explanatory variables (see Table 5.1 below) to be fitted to a logistic regression model in Section 5.2.

Table 5.1 Selected variables, their data types, labels, value levels and measure scale

	Short name	Variable label	No. of categories	Type	Measure scale
1	*FSC*	Food consumption score (r.v.)	NA	N	Scale
2	*FSG*	Food consumption score group (r.v.)	3	C	Ordinal
3	*State*	State	10	C	Nominal
4	*HHType*	Household residential type	3	C	Nominal
5	*HHsize*	Number of household members	NA	N	Scale
6	*Edlevel*	Level of education attained	4	C	Nominal
7	*SexHHH*	Sex of household head	2	C	Nominal
8	*OwnLand*	Owned land for agriculture	2	C	Nominal
9	*UseLand*	Used land for agriculture	2	C	Nominal
10	*PlantLand*	Land planted previous year	2	C	Nominal
11	*LStock*	Owned livestock	2	C	Nominal
12	*Migrates*	Usually migrated (better livelihoods)	2	C	Nominal
13	*Harvests*	No. of harvests in one year	2	C	Nominal
14	*FoodLast*	Months food lasted (normal season)	NA	N	Scale
15	*VGarden*	Vegetable garden	2	C	Nominal
16	*LivSource*	Main sources of livelihood	12	C	Nominal
17	*CerTub*	Main source of cereals/tubers eaten	8	C	Nominal
18	*Meals*	Meals per day (adults + children)	NA	N	Scale
19	*FoodShk*	Experienced food shock	2	C	Nominal
20	*FoodAid*	Received food aid	2	C	Nominal
21	*WIQntile*	Wealth index quintiles	5	C	Ordinal

Key: r.v.= response variable; N=Numeric; C=Categorical; NA=not applicable

5.1.1 Exploratory analysis based on food consumption score as a continuous variable

This sub-section aims at inspecting the distribution of the food consumption score (*FCS*) observations from a sample size of 9 220 households. However, the sample includes some 395 (4.3%) missing values as well. Although not really of any significant importance to this study, some readers might want to find out about the distribution of missing values by state. Most

(81%) of the missing values are found in Western Bahr el-Ghazal State (21%), Jonglei (19%), Warrap (15%), Northern Bahr el-Ghazal (15%) and Unity (11%).

The first procedure is to examine common descriptive statistics from the sample as shown in Table 5.2 below. It should be noted that although the point estimates of *range*, *minimum* and *maximum* food consumption score yield awkward values, this has not affected the mean significantly; since the extreme values close to the lower range, i.e., 0.5 to 3.5 food consumption score, arise from 86 cases only (or only 1 *per cent* of the data). The values to the upper end of the distribution of the data, i.e., between 100 to105, amount to only 0.4 *per cent* (37 cases). The fact that the sample is reasonably big means these extreme cases are nothing to worry about. Instead, it implies that the dataset gives hope for a better fitting model, although this result will have to be later confirmed after fitting a model with all the predictors selected.

Table 5.2 Measures of central tendency for the food consumption score variable

Statistic	**Value**
N	8 825[a]
Range	104.5
Minimum	0.5
Maximum	105.0
Mean	40.9
Standard error of mean	0.23
Standard deviation	21.66

[a] Valid cases only (i.e. sample size less missing cases).

The second type of analysis features plotting of the observations using a frequency distribution histogram. The plot (see Figure 5.2) shows that the distribution of food consumption scores tends to normality with a slight skew to the lower end of the scale. This heralds hope to the rest of the analytical process, as there is an indication of a fair distribution of observations of the response variable.

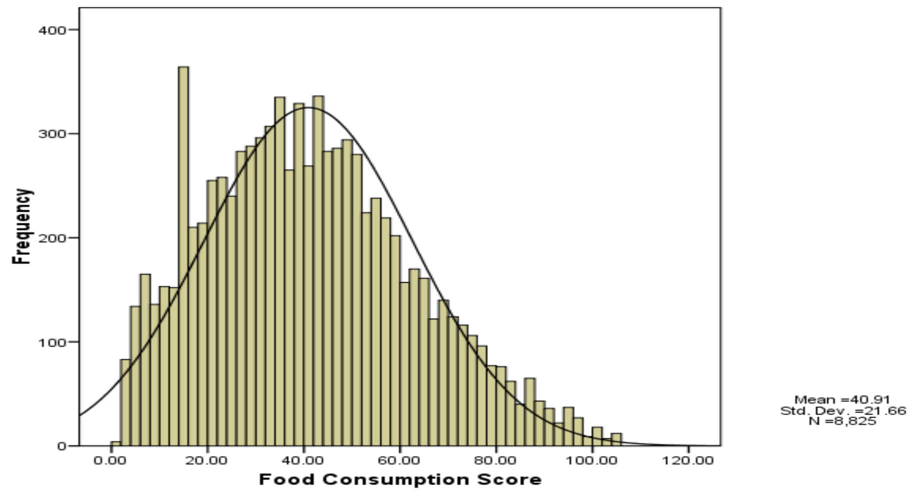

Figure 5.2: Frequency distribution histogram of food consumption score (*FCS*)

Another promising sign of normality in the distribution of the data is using the *P-P Plot*, otherwise known as probability plot. A *P-P Plot* plots the cumulative proportions (observed cumulative probabilities) of a variable against the proportions (expected cumulative probabilities) of any of a number of test distributions. *P-P plots* are generally used to determine whether the distribution of a variable matches a given distribution whereby clustering of points around the straight line indicates the variable matches the test distribution specified (SPSS, 2006). Figure 5.3 assures that the distribution of food consumption scores tends to normality.

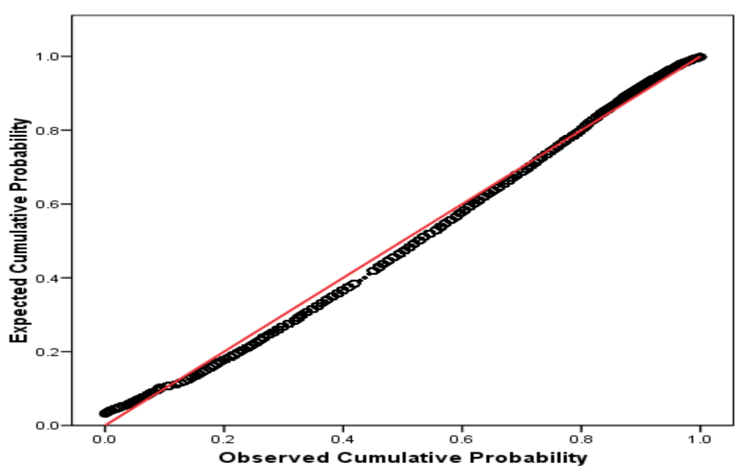

Figure 5.3: Normal P-P Plot of Food Consumption Score

The *Detrended Plot* option of SPSS enables plotting of observed cumulative values against deviations from the expected values. Deviations are calculated by subtracting the expected value from the observed value. As can be observed in Figure 5.4, the distribution of expected probabilities from observed cumulative probability deviations is fairly good, as the points appear to be tightly following a linear distribution and both negative and positive deviations seem to show balanced distribution. It is also observed that the points are densely clustered indicating very little variability. The two types of probability distributions show no visible outliers. In addition, the deviations lie in the interval -0.04 and 0.04, which is very close.

The two illustrations could serve the purpose of showing evidence of closeness of the data to a normal distribution. However, it is worthwhile examining an equally popular method for exploring the distribution of an interval scale variable such as the *FCS*. The simple *Boxplot* method summarises a single numeric variable within categories of another variable (SPSS, 2006).

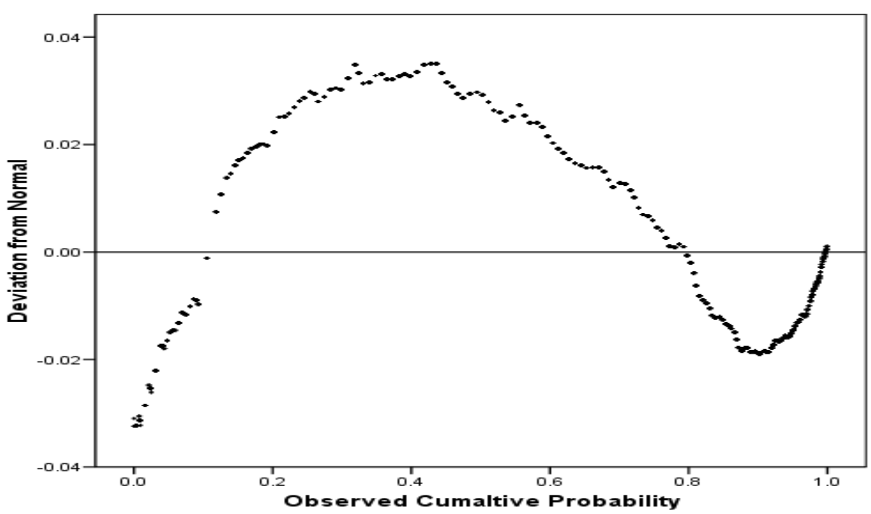

Figure 5.4: Detrended Normal P-P Plot of Food Consumption Score

Boxplots are used in descriptive exploratory analysis to show the median and quartiles as well as outlier and extreme values for a scale variable. The method uses the interqurtile range (the difference between the 75th and 25th percentiles and corresponds to the length of the box. In the boxplot of Figure 5.5, each box shows, the median, quartiles and extreme cases of the food consumption scores within a state. Values between 1.5 and 3 box length from the upper or lower edge of the box are classified as as *outliers*. Values above 3 box length from the upper or lower edge of the box are *extreme*. The length box is the interquartile rage. The boxplot of Figure 5.5 shows that four states (Unity, Western Bahr el-Ghazal, Central Equatoria and Eastern Equatoria) are free of outliers or extreme cases. Three states have one extreme case each and three have between 4 and 6 extreme cases.

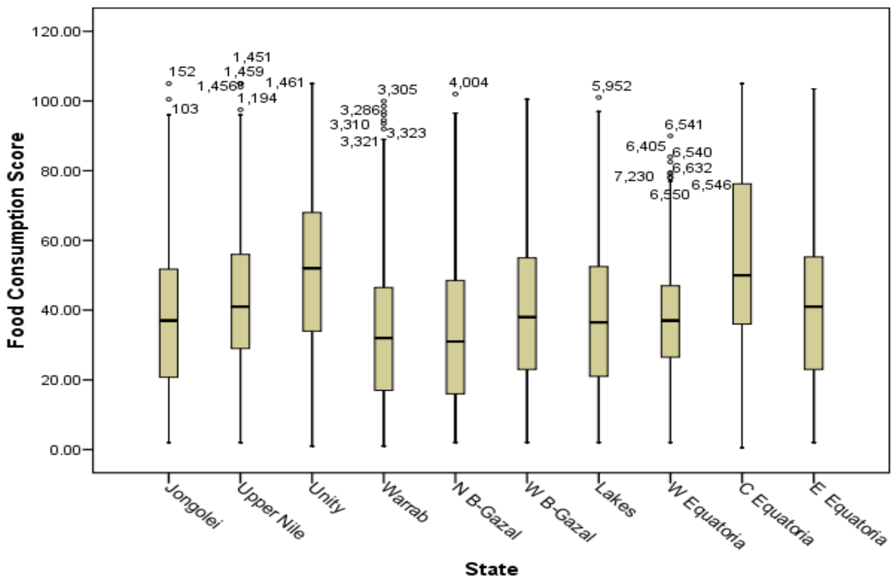

Figure 5.5: Box plots of Food Consumption Score (*FCS*) by state

A boxplot examining the distribution of food consumption scores by food consumption groups, shown in Figure 5.6, reveals 6 extreme values in the '*Good Food Consumption*' group. These extreme cases are shown numbered and will be removed if the examination of the model later reveals evidence of *lack of goodness of fit*.

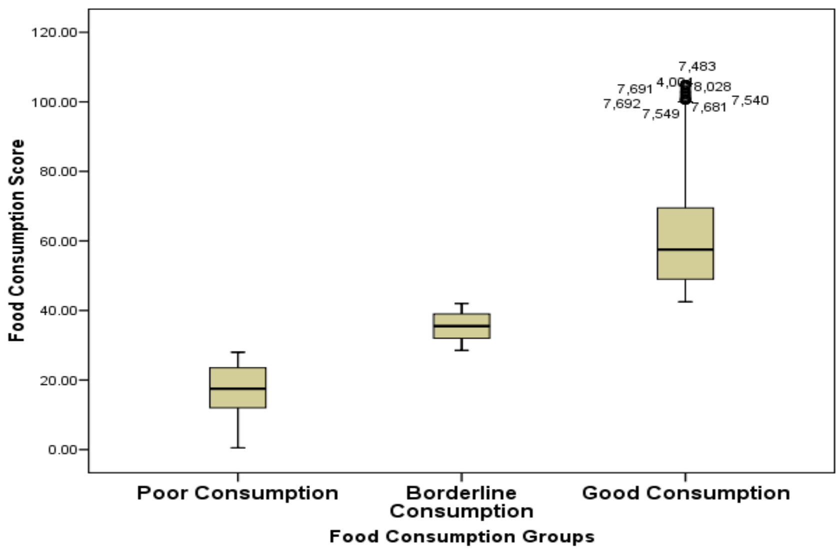

Figure 5.6: Box plot of food consumption score by food consumption group

In general, the four types of exploratory analysis of the distribution of food consumption scores as a numeric variable give hope of a good model although the *boxplot* method distinctly reveals extreme cases. However, the results of the analysis using frequency distribution *histogram*, *P-P Plot* and point measures of central tendency (mean, median and range,) give motivation that the dataset is quite good. Indeed the large sample size must have played a vital role in minimising the effect of the relatively few (21) extreme cases from influencing the model.

5.1.2 Exploratory analysis of linear relationships based on correlation statistics

Exploratory analysis further aims at exploring linear relationships by examining the correlation of each of the three covariates described above with the ungrouped food consumption scores. Correlations are derived by cross-tabulating two continuous variables: dependent and independent. SPSS outputs two types of correlation coefficients and their significance values under selected level of significance. These are: *Pearson's* and *Spearman's* correlations. Another approach which does not apply cross-tabulation and which in SPSS means selecting the *Correlation* command in the *Analyze* menu, gives only significant values of the two methods but does not give the correlation statistics. Table 5.3 displays Pearson's R and Spearman's correlation coefficients for each of the three covariates together with their respective significance probabilities under the 0.01 significance level (*2-tailed* test).

Table 5.3 Correlation statistics between each of three covariates and food consumption score

Food Consumption Score × ...	**Pearson's R**		**Spearman's Correlation**	
	Value	**Significance**[a]	**Value**	**Significance**[a]
Number of household members	0.001	0.933*ns*	-0.025	0.044 *sig*
Months of food availability post harvest	0.157	0.000 *sig*	0.169	0.000 *sig*
Number of meals per day	0.079	0.000 *sig*	0.098	0.000 *sig*

Notes: [a] Significance is based on *2-tailed* test; *sig* = significant; *ns* = not significant

Results of testing for significance of correlations between food consumption score and each of the three covariates vary between the two types of tests. Whereas Pearson's R indicates a non-significant value under the 0.01 significance level for the first covariate, household size, the Spearman's test is totally different. Again, Spearman's statistic yields a negative correlation in total contrast to the Pearson's test, which gives a positive value. Based on Pearson's correlation coefficients, the results show that each of the three covariates has significant positive correlation with food consumption scores except for household size, which is non-significant under the 0.01 significance level. Spearman's test shows significant probabilities for all relations involving the three covariates. These results will later (in the next subsection) be compared with results of relationships when the three variables are grouped and their categories are tested with the ordinal categorical values of food consumption scores.

5.1.3 Exploratory analysis based on food consumption score as a discrete ordinal variable

In this section, the independent variable food consumption scores is treated as a discrete categorical variable. Relationships between two categorical variables, dependent and independent, can be explored by cross-tabulating them, comparing the percentages (or counts) and determining their significance based on Pearson's Chi-Square (χ^2) statistic. Discussing 19 cross-tabulations for each predictor by response variable is a long process. Therefore, only a summary of the variables, Chi-Square statistic, Likelihood Ratio test statistic, their *p*-values and degrees of freedom are displayed as in Table 5.4 below.

In order to generate more reliable statistics, each of the three covariates (numerical predictors) has been grouped into appropriate categories. Leaving the variable ungrouped might result in cells with zero values and hence unreliable statistics. Number of household members has been classified into 4 groups: 1 to 3, 4 to 6, 7 to 9 and 10 members and more. Months of food availability from harvest is a variable that positively correlate with food consumption group as shown in Table 5.3 above. The dataset reveals that Months of food supply range from 0 to 35. Four groups of months of harvest food supply have been created, which are: 0 to 3 months, 4 to 6 months, 7 to 10 months and 11 months and above. Lastly, the number of meals eaten per day by both adults and children has been grouped into 4 intervals: 1, 2, 3 and 4 meals and more.

All statistical tests under the *Likelihood Ratio Test* are aimed at testing the *null hypotheses* that there is no difference between levels of predictors in their association with the three ordered categories of the food consumptions group. *Alternative hypotheses state* that all levels of each predictor differ significantly in their association with the three categories of food consumption scores. For the Pearson's Chi-Square technique, tests are aimed at testing the *null hypotheses* suggesting that there is no difference between observed and expected values of the combinations of levels of predictors and categories of the response variable. In other words, null hypotheses assume that any differences observed between levels of values are due to chance alone.

Test results owing to Pearson's Chi-Square and Likelihood Ratio Test techniques reveal very highly significant probabilities for 16 of the nineteen variables and highly significant differences for number of household members. Only two variables, sex and level of education of household heads, show no significant results. Hence, according to these procedures, there is no adequate statistical evidence to suggest that there was difference between male and female household heads in food consumption of households. However, the rest of the variables are statistically determined to be factors and covariates that very highly determine the way people consumed food.

Table 5.4 Summary statistics from two dimensional cross-tabulations of food consumption score and each of the explanatory variables

	Variable label	**Pearson's Chi-sq**		**LR Test**		**DF**
		Value	**Significance**	**Value**	**Significance**	
1	State	609.78	0.000	628.31	0.000	18
2	Household residential status	26.51	0.000	26.25	0.000	4
3	Number of household members	15.26	0.018	15.28	0.018	6
4	Level of education attained	**9.32**	**0.156**	**9.37**	**0.154**	**6**
5	Sex of household head	**1.59**	**0.451**	**1.58**	**0.454**	**2**
6	Owned land for agriculture	45.91	0.000	46.54	0.000	2
7	Used land for agriculture	22.98	0.000	23.62	0.000	2
8	Land planted previous year	77.67	0.000	78.34	0.000	2
9	Owned livestock	313.50	0.000	314.95	0.000	2
10	Usually migrated	109.35	0.000	109.22	0.000	2
11	Number of harvests in one year	111.42	0.000	116.02	0.000	2
12	Months food lasted (normal season)	540.57	0.000	139.17	0.000	6
13	Vegetable garden	108.52	0.000	110.54	0.000	2
14	Main sources of livelihood	395.37	0.000	388.10	0.000	22
15	Main source of cereals/tubers eaten	56.68	0.000	55.32	0.000	14
16	Meals per day (adults + children)	386.31	0.000	371.32	0.000	6
17	Experienced food shock	11.01	0.004	10.97	0.004	2
18	Received food aid	14.64	0.001	14.73	0.001	2
19	Wealth index quintiles	69.27	0.000	69.08	0.000	8

Note: Chi-Sq= Pearson's Chi-Square; LR=Likelihood Ratio; DF=Degrees of Freedom

5.1.4 Conclusion

The exploratory procedures above give sufficient evidence and hence motivation to proceed with fitting a multiple logistic regression model for the food consumption data. So far statistics lead to a general conclusion that the response variable food consumption scores or group is a function of at least fifteen factors. These factors can be ascertained to be determinants of food security or insecurity at the household level. Based on this analysis, further implications of the results motivate policy focussed on advocacy regarding households' commitment to utilise opportunities availed to them. It is of interest to note that states differed highly significantly in food consumption. Further analysis based on the robust logistic regression techniques will be able to reveal which states were better in food consumption and which ones were worse. Same investigation of disparities between levels (or groups) of variables will be examined using the Ordinal Logistic Regression model. As in Ordinary Linear Regression, the Logistic Regression enables fitting all the variables in one model rather than carrying out separate analyses. This will then enable determination of whether each of the variables, when taken together, still yield significant results of a relationship.

5.2 Logistic Regression analysis based on the Proportional Odds Model

In this section analysis makes use of the logistic regression technique based on the Proportional Odds Model. As stated in Chapter 3, logistic regression is a member of a family of Generalized Linear Models, a methodology developed by McCullagh (1980, McCullagh and Nelder, 1989), which uses a generalisation of the linear regression for prediction of the cumulative probabilities of the ordered categories of the response variable. The method enables fitting of a set of equations for each category of the ordered dependent variable with each equation giving predicted probabilities of being in the corresponding category or any of the categories that are lower in rank. An important assumption distinguishing the proportional odds model is that the predicted values of categories of the ordered variable are a set of parallel lines. This then leads to testing for an alternative hypothesis of non-parallel lines.

5.2.1 Choice of a Link Function

The proportional odds model, rather than predicting the actual cumulative probabilities, it predicts a function of their values called the *link function*. SPSS Ordinal Regression Procedure or PLUM provides the choice among several link functions of which the most commonly used are *logit* and the *complementary log-log* link function (SPSS, 2006). The latter is suitable for the ordinal regression and will be the one used in the model. It is advisable, however, that before choosing which link function to use it is worthwhile examining the distribution of the response variable first.

A bar chart distribution of food consumption group values shown in Figure 5.7 shows the *Borderline Consumption* (23.8%) group as the smallest, followed by *Poor Consumption* (*31.5%*)

and *Good Consumption* (44.6%). As the descriptive statistics analysis of Section 5.1 shows that the ungrouped mean of food consumption score is 40.91, values below the mean food consumption score can be regarded as poor or borderline poor. Hence, there is strong reason to look back at the ungrouped or continuous food consumption scores distribution of Figure 5.2 in which it is established that there is slight skewness towards the lower values of the frequency distributions and which shows that the values above the mean have lower frequencies.

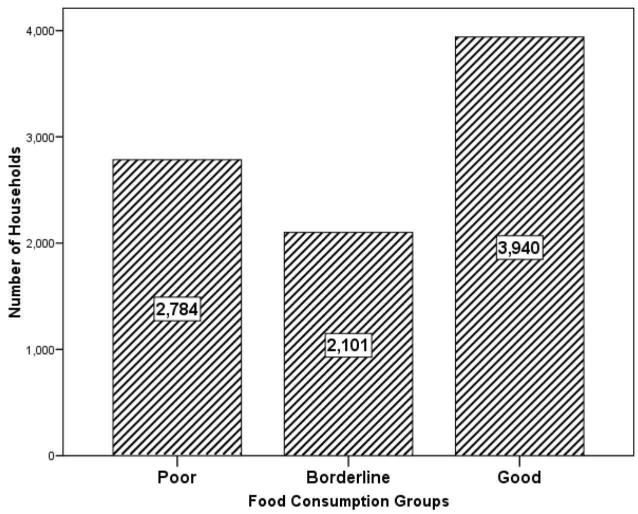

Figure 5.7: Distribution of households by categories of Food Consumption Scores

It is also observed that the ungrouped distribution of the data reveals a very wide range with extreme values of 0.5 and 105 food consumption scores. These extreme values call for a careful choice of the suitable link function. SPSS version 16 avails 5 link functions for the ordinal regression model. A link function transforms the cumulative probabilities for estimation of the model. Table 5.5 outlines the five link functions, their notational formulations and application.

The explanation of the rationale behind choice of the link function leads to a split decision between use of *complementary log-log* or Cauchit link functions. A model with the former will be explored and in case it is found to be unsuitable, then the latter link function will be tried.

Table 5.5 A summary of five link functions used in Ordinal Regression

Function	Form	Typical application
Logit	$\log(\xi/1-\xi)$	Evenly distributed categories
Complementary log-log	$\log(-\log(1-\xi))$	Higher categories more probable
Negative log-log	$-\log(-\log(\xi))$	Lower categories more probable
Probit	$\phi^{-1}(\xi)$	Latent variable is normally distributed
Cauchit (inverse Cauchy)	$\tan(\xi-0.5))$	Latent variable has many extreme values

Source: SPSS version 16 (2006)

5.2.2 Fitting the ordinal logistic regression to the food consumption data

In this study the aim behind fitting an ordinal logistic regression model is to predict the ordinal outcome of the food consumption scores that has three categories: *poor, borderline* and *good consumption*. The model will fit all of the predictors including sixteen factors and three covariates (see Table 5.1 above). That means the results of the exploratory analysis that found two predictor variables as not being important, i.e. not giving statistically significant difference between their levels, are ignored in the initial fitting of the model. Where the model will confirm these variables as non-significant in predicting the outcome, they will be dropped and the model re-fitted.

5.2.3 Running the analysis

To run the ordinal regression model from the SPSS menus, the steps are: *Analyze, Regression, Ordinal...* This leads to Ordinal Regression Dialogue with many options to choose from and specifying values for the model. The *location-only* ordinal regression model will be fitted first and in case there will be evidence that this model is inadequate for the data, the alternative *scale component* model will be fitted to the data. In specifying options for the model, the link function chosen will be the *Complementary Log-Log* and 95% confidence interval. The output options selected will be as displayed in Figure 5.8. The SPSS PLUM procedure allows building of a model, generating predictions and evaluating the importance of various predictors, where the dependent variable is ordered categorical. The PLUM code is included in Appendix 3 for reference. However, in this section menu options were followed.

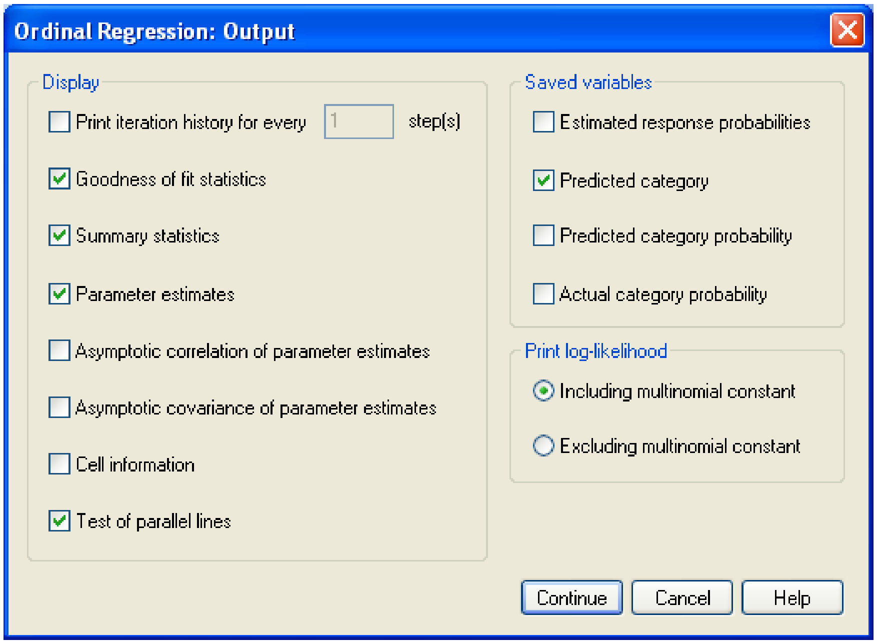

Figure 5.8: Selected options for the output of fitting Ordinal Logistic Regression Model

5.2.4 Evaluating the model

After running the specified model, SPSS generated an output that displays the warning message shown in Figure 5.9 right at the top of the output. The warning message informs the user about the number of cells generated that contain zero frequencies. It is apparent that the three continuous covariates must have generated the cells with zero frequencies. Note that all the categorical and the continuous variables transformed into categorical variables in Subsection 5.1.3 did not generate zero frequencies. That means certain values of the individual covariates when cross-matched with food consumption scores did not show frequencies with zero values. It is quite normal that, with the huge number of cases (households) of each of the three covariates cross-tabulated by three categories of the response variable (food consumption score) and by the rest of the 18 other predictors, combine to form empty cells. In the case of this study, the number and percentage of cases is two thirds. Other datasets would generate even higher percentage of cells with zero values.

Warnings

There are 5630 (66.6%) cells (i.e., dependent variable levels by combinations of predictor variable values) with zero frequencies.

Figure 5.9 SPSS output message alert about presence of cells with zero values

A model generating a very large number of empty cells should lead to taking caution in interpreting the results of the fit statistics such as Chi-Square statistics. If the model included a few factors, say, four to six, it would have been recommended to display information about individual cells by selecting the Cell Information checkbox of the Ordinal Regression Outputs dialogue shown in Figure 5.8 above. However, this option is not advised for a very large number of variables as well as models with continuous variables or factors with many levels, such as in the case here. Doing this will result in very large multidimensional tables and which often result in long processing time when the model is run.

SPSS outputs characteristically display Case Processing Summary Information. This table is usually ignored because it only informs the user of the percentage of cases of each level of a variable selected for analysis or included in the model. If it serves any meaningful purpose, the table shows how certain individual levels of a variable have bigger proportion of households than others and to indicate how certain levels end up with negative or positive coefficients. The Case Processing Summary Table will not be displayed here.

A good strategy in evaluating a logistic regression model is to determine whether the model gives adequate predictions well before examining individual values of predictors. This is done by examining the Model-Fitting Information displayed in Table 5.6 as the third set of information displayed in the SPSS output.

Table 5.6 SPSS output of model fitting information

Mode l	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	6025.082			

Final	5413.788	611.294	49	.000

Link function: Complementary Log-log.

Table 5.6 shows the values of the *-2 Log-Likelihood* statistic calculated after fitting a model with the *intercept* only (i.e. a model that does not include any predictor variable) as well as the *final model*, which in other texts is known as the *saturated model* (to which all the predictors were fitted). The difference between the two *Log-Likelihoods* yields the Chi-Square statistics (McCullagh and Nelder, 1989), which is shown as 611.294. Looking up the Chi-Square value in statistics significance tables, constructed from 95% confidence level and 49 degrees of freedom (df), the model is shown to have significant difference between the final model, with all predictors included and that with only the intercept fitted. This is an indication of an improvement of the model with predictors over that without predictors. It therefore leads to the conclusion that the fitted model gives better predictions than if interpretations were based on marginal probabilities of the categories of the response variable food consumption groups.

The output also displays the *Goodness-of-Fit* table for the model based on two hypothesis test statistics, namely; *Pearson's Chi-Square* and another Chi-Square statistic based on the *Deviance* (see Section 3.2.6). Both statistics are aimed at testing whether the observed data are inconsistent with the fitted model. Large significance values lead to a conclusion that both the observed values and the values predicting the model are similar and that the model is good. This is indeed a very good and confidence building outcome about the model. As stated above, a model with too many empty cells renders conclusions about the goodness of a model or one that is not suitable, as misleading. In other words, the goodness of fit of the model, i.e. that it follows a Chi-Square distribution, cannot be ascertained due to the large number of empty cells (see Figure 5.9) and invalidates the accuracy of the statistics and their significance. Table 5.7 shows large significant values compared to the significance level of 0.05 – a result that could have led to the conclusion that the fit of the model is good, had it not been for the limitations described above.

Table 5.7 Goodness-of-Fit statistics for the model.

	Chi-Square	df	Sig.
Pearson	5722.532	5581	.091

Deviance	5411.016	5581	.947

Link function: Complementary Log-log.

The next result shows the *Coefficient of Determination* or R^2 statistic. Various texts on linear regression models describe the R^2 (*see* Menard, 2002; Hosmer & Lemeshow, 2000). Basically the R^2 is said to summarise the proportion of variance in the dependent variable associated with the independent variables. Larger R^2 values (i.e. closer to 1) indicate that more of the variation is explained by the model. However, in models with categorical dependent variables, computation of a single R^2 statistic, that has all the characteristic of an R^2 in a linear regression model, is not possible. Instead, SPSS computes a semblance of R^2 statistics for the ordinal logistic regression known as *Pseudo-R^2*. SPSS includes computations of the *Pseudo-R^2* according to Cox and Snell (1989), Nagelkerke (1991) and McFadden (1991). The statistic is based on the *Log-Likelihood* for the *final* model compared with the *Log-Likelihood* of the *baseline* model (please see the given references for the description of the rest of the procedures). The *Pseudo-R^2* values shown in Table 5.8, although not too small considering the inclusion of the three interval scale variables in the fitted model, give some reason for revision of the model in order to generate better predictions (SPSS, 2006).

Table 5.8 Pseudo R^2 values

Cox and Snell	.195
Nagelkerke	.221
McFadden	.101

Link function: Complementary Log-log.

The next step in evaluation of the model is to examine its generated predictions. Note that when specifying the options of the Ordinal Regression Outputs in the SPSS dialogue displayed in Figure 5.8, the option “Predicted category” in the Saved Variables group, is selected. This enables SPSS to generate another variable *PRE-1* which has predicted values of food consumption groups. Since interest is centred on generating correct predicted categories of the response variable based on the values of the predictor variables, for determining the performance of the fitted model, *Classification Tables* are constructed by cross-tabulating the predicted categories with the observed (or actual) categories. Table 5.9 contains the frequencies and percentages of predicted and observed response categories.

Table 5.9 Classification table of predicted by observed categories of food consumption groups

			Predicted Response Category		
			Poor Consumption	Good Consumption	Total
Food Consumption Groups	Poor Consumption	Count	497	381	878
		% within Food Consumption Groups	56.6%	43.4%	100.0%
	Borderline Consumption	Count	272	415	687
		% within Food Consumption Groups	39.6%	60.4%	100.0%
	Good Consumption	Count	210	1049	1259
		% within Food Consumption Groups	16.7%	83.3%	100.0%
Total		Count	979	1845	2824
		% within Food Consumption Groups	34.7%	65.3%	100.0%

The model appears to be doing fairly well in predicting the response categories at least for the most frequent category *Good Consumption* where it correctly classifies 83.3% of the cases. On the other hand, the model has done marginally fairly in classifying cases of the *Poor Consumption* category, where it only classifies 56.6% of the cases as correct. Cases in the *Borderline Consumption* category are more likely to be classified in the *Good Consumption* category than in the *Poor Consumption* category. There is indication of some slight problem in the way the ordinal response scale was defined, especially with regards to how the second interval *Borderline Consumption* was defined. This outcome, therefore, calls for redefining (recoding) the categories of the ordinal response in order to improve it. Indeed examination of the distribution of the response variable food consumption group (FCG), which is shown in Figure 5.7, clearly shows that the middle category of the response variable food consumption score is the smallest of the three categories. This finding will help in analysis of data from future household food security surveys (by improving the intervals of the food consumption scores) based on more or less fairly distributed frequencies of the variable categories.

5.2.5 Test of parallel lines

The test of parallel lines tests the proportional odds assumption. Alternatively, the test is used for testing the assumption that model parameters (estimates of slope coefficients or $\hat{\beta}_j$ values) are the same for all categories of the response variable. Basically, the test enables comparison of the estimated model with one set of the model coefficients for all categories to a model with a separate set of coefficients for each category (SPSS, 2006; Hosmer and Lemeshow, 2000). In other words, it is assumed that with the same slope throughout the categories of food consumption groups and different intercepts, log-linear equations of the relationship of an explanatory variable with food consumption groups, if plotted on a plane with the *logits* of *FCS* as the dependent variable and levels of a factor as the independent variable, there will be parallel lines.

As stated in Table 5.10, the test model tests the *null hypothesis* H_0: all slope coefficients $\hat{\beta}_j$'s, also known as *location parameters*, are the same across the categories of the food consumption scores (i.e. $\hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_k$).

Table 5.10 SPSS output of the Test of Parallel Lines [c]

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	5413.788			
General	5291.444(a)	122.344(b)	49	.000

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.
a The log-likelihood value cannot be further increased after maximum number of step-halving.
b The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.
c Link function: Complementary Log-log.

The highly significant test of parallel lines indicates that the null hypothesis is to be rejected. That is, the coefficients differ a lot across the categories of food consumption scores so much so that no two lines of the same slope, for different categories, can be parallel. This fit problem is obviously due to improper ordering of the categories of the response variable. As suggested in 5.2.3, the middle category of *FCS borderline consumption* will need to be widened to include values from the category *good consumption*, which must have been “*over-lumped*”. Therefore, there is adequate evidence that the model needs to be refitted.

5.2.6 Interpreting the model

The influence of each predictor variable on the response variable is determined by examining the coefficients of each factor or covariate. Interpretation of values of coefficients differs between factors and covariates. For covariates, positive coefficients indicate positive relationships between predictors and outcomes. On the other hand, negative coefficients indicate negative relationships. A covariate with an increasing positive value of a coefficient corresponds to an increasing probability of being in one of the lower level categories of the cumulative response. For factors, a factor level with a greater coefficient indicates a greater probability of being in one of the lower level categories of the cumulative response. A factor with a negative sign (indicator) indicates that its level has negative effect on the corresponding category of the response variable. The converse is true for a factor with a positive sign. Interpretation of the model can, therefore, be based on the parameter estimates.

As shown in Table 5.11 below, SPSS tabulates parameter estimates of the *logit* coefficients of the model, gives lower and upper limits of the 95% confidence intervals, displays values of the Wald Test, degrees of freedom (df) and the significance probability (sig.). Interest centres on knowing whether an estimate is significant or not in order to decide whether to reject or not to reject the null hypothesis. The null hypothesis states that each parameter estimate contributes zero to the relationship with the response variable. For an estimate to differ significantly from zero, its significance value must be less than the significance level of 0.05. Therefore, according to the model, 12 variables (including all the 3 covariates and 9 factor variables) have levels with significant values compared to reference levels (in case of factors).

Table 5.11 Edited SPSS output of parameter estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[FCG = 1]	1.648	.524	9.878	1	.002	.620	2.676
	[FCG = 2]	2.528	.525	23.176	1	.000	1.499	3.557
Location	HHsize	.023	.010	5.886	1	.015	.005	.042
	FoodLast	.049	.012	17.248	1	.000	.026	.072
	Meals	.297	.044	46.463	1	.000	.212	.383
	[State=71]	.721	.139	26.998	1	.000	.449	.993
	[State=72]	.739	.136	29.423	1	.000	.472	1.006
	[State=73]	1.123	.145	60.279	1	.000	.840	1.407
	[State=81]	.397	.126	9.986	1	.002	.151	.643
	[State=82]	.399	.122	10.660	1	.001	.159	.638
	[State=83]	.622	.123	25.488	1	.000	.380	.863
	[State=84]	.361	.108	11.194	1	.001	.150	.573
	[State=91]	.236	.107	4.829	1	.028	.025	.446
	[State=92]	1.184	.124	91.397	1	.000	.941	1.426
	[State=93]	0(a)	.	.	0	.	.	.
	[HHType=1]	.142	.100	2.003	1	.157	-.055	.339
	[HHType=2]	.080	.113	.494	1	.482	-.143	.302
	[HHType=3]	0(a)	.	.	0	.	.	.
	[Edlevel=1]	.037	.094	.154	1	.695	-.148	.222
	[Edlevel=2]	.057	.098	.337	1	.562	-.135	.248
	[Edlevel=3]	-.154	.109	1.980	1	.159	-.368	.060
	[Edlevel=4]	0(a)	.	.	0	.	.	.695
	[SexHHH=1]	-.029	.078	.139	1	.709	-.182	.562
	[SexHHH=2]	0(a)	.	.	0	.	.	.159
	[OwnLand=1]	-.166	.081	4.215	1	.040	-.324	-.008
	[OwnLand=2]	0(a)	.	.	0	.	.	.
	[UseLand=1]	-.381	.118	10.468	1	.001	-.611	-.150
	[UseLand=2]	0(a)	.	.	0	.	.	.
	[PlantLand=1]	.143	.062	5.344	1	.021	.022	.264
	[PlantLand=2]	0(a)	.	.	0	.	.	.
	[LStock=1]	.286	.059	23.310	1	.000	.170	.402
	[LStock=2]	0(a)	.	.	0	.	.	.
	[Migrates=1]	.117	.065	3.274	1	.070	-.010	.244
	[Migrates=2]	0(a)	.	.	0	.	.	.
	[Harvests=1]	-.088	.076	1.346	1	.246	-.236	.060
	[Harvests=2]	0(a)	.	.	0	.	.	.
	[VGarden=1]	.364	.058	40.068	1	.000	.251	.477
	[VGarden=2]	0(a)	.	.	0	.	.	.

Table 5.11 (Continued...)

	Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
[LivSource=1]	**.628**	**.193**	**10.642**	**1**	**.001**	.251	1.006
[LivSource=2]	**.358**	**.191**	**3.499**	**1**	**.061**	-.017	.733
[LivSource=3]	**.479**	**.237**	**4.082**	**1**	**.043**	.014	.943
[LivSource=4]	-.042	.239	.031	1	.860	-.511	.426
[LivSource=5]	.167	.246	.460	1	.498	-.316	.650
[LivSource=6]	-.143	.208	.473	1	.492	-.551	.265
[LivSource=7]	-.093	.233	.158	1	.691	-.549	.364
[LivSource=8]	-.129	.303	.183	1	.669	-.722	.464
[LivSource=9]	.444	.369	1.448	1	.229	-.279	1.168
[LivSource=10]	.422	.315	1.790	1	.181	-.196	1.040
[LivSource=11]	-.044	.236	.035	1	.852	-.507	.418
[LivSource=12]	0(a)	.	.	0	.	.	.
[CerealSrce =1]	.571	.439	1.692	1	.193	-.289	1.431
[CerealSrce =2]	.641	.442	2.105	1	.147	-.225	1.507
[CerealSrce =3]	.725	.506	2.057	1	.152	-.266	1.717
[CerealSrce=4]	-.192	.513	.140	1	.708	-1.197	.814
[CerealSrce =5]	-.086	.718	.014	1	.904	-1.492	1.320
[CerealSrce =6]	**1.203**	**.586**	**4.217**	**1**	**.040**	.055	2.350
[CerealSrce =7]	.551	.450	1.495	1	.222	-.332	1.434
[CerealSrce =8]	0(a)	.	.	0	.	.	.
[FoodShk=1]	.077	.055	1.983	1	.159	-.030	.185
[FoodShk=2]	0(a)	.	.	0	.	.	.
[FoodAid=1]	-.026	.064	.163	1	.686	-.152	.100
[FoodAid=2]	0(a)	.	.	0	.	.	.
[WIQntile=1]	.150	.094	2.546	1	.111	-.034	.335
[WIQntile=2]	**.232**	**.088**	**7.014**	**1**	**.008**	.060	.404
[WIQntile=3]	.131	.082	2.570	1	.109	-.029	.292
[WIQntile=4]	.041	.077	.285	1	.593	-.110	.192
[WIQntile=5]	0(a)	.	.	0	.	.	.

As revealed by the exploratory analysis using 2-way cross-tabulations and Chi-Square statistics, level of education and sex of household head are found to be highly non-significant based on the 0.05 significance level. In addition to these two, the model determined as non-significant the following: (i) Residential type of household (i.e. *resident, internally displaced* and *returnee*) (ii) Occasional migration or transient nature of household; (iii) Number of harvests in a year; (iv) Livelihood source being other than production and dependence on natural resources; (vi) Source of cereals (sorghum and millet) being other than gifts from relatives and; (vii) Experience of food shock; (viii) receiving of food aid and; (ix) wealth index quintile being other than *poorer*. The factors determined to be highly non-significant will be dropped from the model and the model re-evaluated. The factors that are marginally significant or that have significant levels will be included together with those shown to be highly and very highly significant.

5.2.7 Revising the model

Results from the above analysis clearly call for fitting a revised model with a *Cauchit* link function. The revised model yields results as displayed in Table 5.12, which shows an improvement in fitting a model with *Cauchit* link function over one with the *Complementary Log-Log* as the link function.

Table 5.12 Comparison between results of two models with different *link functions*

		Model No. 1[a]		**Model No. 2**[b]	
Procedure	**Statistic/Variable**	**Estimate**	**Sig.**	**Estimate**	**Sig.**
Model fitting information	Chi-Square	609.984	0.000	592.359	0.000
Goodness-of-fit	Pearson's Chi-Square	5764.351	0.067	5552.447	0.608
	Deviance Chi-Square	5437.738	0.944	5429.950	0.925
Pseudo R-Square	Cox and Snell	0.194	-	0.189	-
	Nagelkerke	0.219	-	0.215	-
	McFadden	0.101	-	0.098	-
Test of parallel lines	Chi-Square	125.925	0.000	125.630	0.0000
Parameter estimates	Household size	0.023	0.017	0.016	0.1744
	Months of harvest food	0.050	0.000	0.091	0.0000
	Number of daily meals	0.298	0.000	0.345	0.0000
	Central Equatoria State [c]	1.145	0.000	1.381	0.0000
	Land ownership	-0.173	0.031	-0.190	0.0603
	Land use	-0.394	0.001	-0.546	0.0002
	Land planting	0.145	0.018	0.218	0.0066
	Livestock ownership	0.219	0.000	0.385	0.0000

Table 5.12 (Continued...)

		Model No. 1[a]		Model No. 2[b]	
		Estimate	Sig.	Estimate	Sig.
Parameter estimates	Household status = 1 (*resident*)	0.142	0.157	0.074	0.565
	Household status=2 (*Internally displaced*)	0.080	0.482	-0.052	0.724
	Education level=1 (*none*)	0.037	0.695	0.002	0.986
	Education level=2 (*primary*)	0.057	0.562	0.002	0.984
	Education level=3 (*secondary*)	-0.154	0.159	-0.231	0.102
	Sex of household head = 1 (*male*)	-0.029	0.709	-0.041	0.681
	Household migrating	0.120	0.062	0.241	0.0026
	Ownership of vegetable garden	0.364	0.000	0.433	0.0000
	Number of times harvested per year	-0.088	0.246	-0.027	0.7763
	Livelihood sources = 1 (*livestock raring*) [d]	0.619	0.001	1.139	0.0000
	Maize and sorghum from relatives' gifts	1.203	0.040	1.843	0.0180
	Experienced shock	0.077	0.159	0.090	0.1992
	Received food aid	-0.026	0.686	-0.017	0.8336
	Wealth index quintile = 1 (*poorest*)	0.152	0.105	0.105	0.3861
	Wealth index quintile = 2 (*poorer*)	0.235	0.007	0.191	0.0863
	Wealth index quintile = 3 (*moderate*)	0.128	0.117	0.126	0.2353
	Wealth index quintile = 4 (*richer*)	0.040	0.598	0.050	0.6201

[a] Model with *Complementary Log-Log* Link Function; [b] Model with *Cauchit* Link Function; [c] Although many states are significant compared to Eastern Equatoria State, only Central Equatoria State has been included for the purpose of comparison; [d] Although all source of livelihoods with more than one level are significant, only the first level is selected for comparison with the last level "other" source.

After the *Cauchit* link function, the second fitted model improved considerably. In fact, the new model has resulted in marked improvement in the estimates and significance values. In the first place, it is observed that although Model Fitting Information shows some small reduction in value of Log-Likelihood Ratio statistics after fitting the final model and a model with only the response variable from a Chi-Square value of 609.98 to 592.359, compared to that of the first model, it still gives a highly significant probability. This very highly significant statistic indicates that the final model gives better predictions of the response variable than analysis based on proportions.

Some really impressive results in examining the new model are the Goodness-of-Fit statistics. Both Pearson's Chi-Square and Deviance Chi-Square values are not significant, which indicates that the data and the model predictions are similar and hence the second model passes the Goodness-of-Fit test, despite the presence of many empty cells as a result of including the three covariates: *household size*, *number of months during which harvest food lasted* and *number of daily meals eaten.* The Pseudo R-Square, although showing lower values than the first model, may be ignored. There is an increase in the Chi-Square value of the Test of Parallel Lines. However, this has not affected the significance of the difference in the levels of predictors associated with categories of food consumption. Interpretations are similar to those of the earlier result.

5.2.8 Classification Table of the final model

Finally, there is need to examine the Classification Table after fitting the second Ordinal Regression model. Table 5.13 examines the predictions generated by the model as an additional variable in the dataset. As interest is centred on correctly predicted categories of the response variable, the model seems to do well in correctly predicting the *good consumption group* (80.5%). However, it does poorly in classifying correctly the lower consumption categories, especially the *borderline* food consumption category. Most (59.1%) of the *borderline* category cases are classified as *good consumption*. The model classifies correctly 57.1 *per cent* of the *poor consumption* cases. Overall, the model Classification Table shows a marked improved of the Cauchit model over the Complementary Log-Log model. However, the two models seem to demonstrate the need to re-scale the ordinal categories of food consumption scores. The ordering criteria was adopted from a guide by the United Nations World Food Programme Vulnerability Analysis and Mapping (VAM) Unit on calculation and use of the food consumption scores in food security analysis" (WFP-VAM, 2008). There is reason to believe that the *borderline consumption* category is too narrow for the wide range of food consumption scores in the dataset.

Some WFP publications use the categories: 0-21 for *poor food consumption*; 21.5-35 for *borderline food consumption* and; >35.5 for *good food consumption* (WFP-CFSVA, 2007).

Table 5.13 Classification table of predicted by observed categories

			Predicted Response Category			Total
			Poor Consumption	Borderline Consumption	Good Consumption	
Food Consumption Groups	Poor Consumption	Count	501	28	349	878
		% within Food Consumption Groups	57.1%	3.2%	39.7%	100.0%
	Borderline Consumption	Count	252	29	406	687
		% within Food Consumption Groups	36.7%	4.2%	59.1%	100.0%
	Good Consumption	Count	218	27	1014	1259
		% within Food Consumption Groups	17.3%	2.1%	80.5%	100.0%
Total		Count	971	84	1769	2824
		% within Food Consumption Groups	34.4%	3.0%	62.6%	100.0%

5.2.9 Results and discussion

The next step is to examine model estimates from fitting the second model. The most glaring change is that two seemingly important variables that were significant in the first model, i.e. size of household and land ownership, were determined by fitting the second model as not significant. Possible interpretation of this finding is that *household size* and *farmland ownership* are not related to *food consumption*, a result contrary to expectations and guesses. A contrary finding is that incidence of *households migrating* from one location to another was marginally not significant (*p*-value = 0.06) in fitting the first model with more predictors and *Complementary Log-Log* link function, but in the second model this predictor is shown as very highly significant (p-value=0.0003)! Households that migrated from place to place in search of pasture or work-related movement of settlement during the year, were *exp(0.241)* = 1.27 time those that did not, after fitting the second model. In fitting the first model, the odds of a household migrating attaining lower food consumption score level were only exp(0.117) = 1.12 times those of households that did not migrate during the year. There was significant difference (*p*-value=0.018) between the source of the cereals sorghum and millet received from relatives as gifts than “other” sources with regard to the probability of being in *poor* or *borderline poor* food consumption groups with the odds of 6 to 1. This result is not contrary to any expectation because dependence on gift food is a sort of coping strategy.

The rest of the predictors show significant results of tests using the *Wald Statistic* in both fitted models. “*Months of food from harvest eaten”* is also positively related to food consumption groups. This is expected as the longer the food from harvest lasts, the greater the probability of being in *poor* or *borderline* food consumption. Also as expected, *number of daily meals eaten* by a household is related positively to food consumption group. The probability of a household being in a *poor* or *borderline* food consumption is improved as daily meals eaten increase.

Households differed significantly between states of residence compared to a reference state (Eastern Equatoria) in their food consumption levels. The dataset shows marked difference between Eastern Equatoria State and the nine other states, namely. The odds of a household in Central Equatoria State being in lower category of food consumption were *exp(1.381)=3.98* or about four times those of a household in Eastern Equatoria State, when the rest of the fitted variables took zero coefficients or not taken into account. The state showing huge disparity compared to Eastern Equatoria State was Unity, with odds of getting a lower food consumption classification of *exp(1.482)=4.4* times those of Eastern Equatoria State while taking other fitted variables as non-existent.

Households that used land for farming prior to the survey were negatively related to the probability of being in the food consumption group *poor* or *borderline*. The odds of a household that used farmland were only *exp(-0.546)*=0.6 times less of being *poor* or in *borderline* food consumption group compared to those of a household that did not use their farmland. In other words, there were more households not using land that risked being in lower food consumption groups. As for the *planting of farmland*, there was significant relationship with food consumption groups. The odds of a household that planted its land were *exp(0.218)=1.2* of being in *poor* or *borderline poor* food consumption group. This means, contrary to expectations, land planting did not improve the household’s food consumption levels. Instead those that did not plant did better to some extent. It is quite likely that the post conflict nature of South Sudan is to blame in that there were more people who lived from sources other than farming. It is possible that a sizeable number of households that planted land had not yet harvested or did not get satisfactory harvest due to poor yield. Other factors such as poor rainfall and dependence on food aid, despite availability of land, could reduce the probability of attaining poor food consumption level. It is noteworthy that *ownership of home gardens* statistically yields highly significant and positive relationship with food consumption groups. The odds of households that owned home gardens or vegetable plots were *exp(0.433)=1.5* times having more probability of falling into the poor or borderline food consumption categories than those that did not own home gardens. In other words, not having a garden increased the probability of being in better food consumption group – not quite what one would expect. Dependence on home gardens would have meant getting either less dietary diversity or smaller frequency of eating a food item high in nutritional value.

It is also interesting to note that *ownership of livestock* was positively related to food consumption score levels. In other words, the odds of a household owning livestock being in a *poor* or *borderline* food consumption group were significantly worse by 1.5 times compared to households that did not own livestock. One would wonder why ownership of important assets like livestock would not improve the chance of eating well. The answer to the question is easy. Most livestock keepers in Southern Sudan did not keep them for meeting their daily dietary requirements or even improving their livelihoods. This finding, however, provokes research into the habits of livestock and common beliefs.

"*Main source of livelihoods*" is an important determinant of dietary consumption and indeed of food insecurity. The software treated the main source of livelihoods "*other*" as a comparator. Analysis reveals that households differed significantly in terms of food consumption. In magnitude terms, livestock rearing, agricultural production, fishing and employment were related to the probability of being in the poor or borderline food consumption groups. The odds of a household which depended on any of these four sources of livelihoods having the risk of getting into a *poor* or *borderline* food consumption category ranged from 2 to 3 times as much as those of a household that used "other" livelihood sources. If these results highlighted anything, it only confirmed the finding on livestock keeping, home gardening and farmland use and the probability of being in poor dietary consumption category (i.e. $FCS \leq 42$).

Also of interest is noting that households *getting cereals and tubers* by exchanging or offering their services (labour) to get the food they ate in the reference period of the survey. Results show that these households were negatively related to lower food consumption categories. In other words, the household that exchanged their services or other items to get cereals or tubers did not fall into the poor or borderline poor food consumption categories. Instead, they had 'favourable' probability of being in a better food consumption group compared to a household getting food from the "*others*" source. In numerical terms the odds of a household exchanging services for food were only $exp(-1.094)$ = 0.33 or a third of those in the "*other*" group; taking other predictors as non-existent. This is interesting because one would expect the livelihoods source "*own production*" or "*market purchase*" to improve the probability of being in the *borderline* or *good* food consumption groups. However, those livelihoods sources did not show any significant difference with the "*other*" source. The post-conflict setting of Southern Sudan could also explain this unexpected result, which might be different under normal conditions.

As regards Wealth Index Quintiles, there was no significant difference between any of its levels as related to food consumption groups. This result contradicts all expectations in that wealth is supposed to be positively related to good food consumption. Moreover, the Pearson's and Likelihood Ratio tests of 2-way relationship based on proportions revealed very strong evidence of relationship of wealth index and food consumption group. However, the modelling enables factors to adjust for one another based on the variance.

5.2.10 Conclusion

In conclusion, it has been shown that the Ordinal Regression model fitted with the *Cauchit* link function tremendously improves the performance of the model and estimation of its parameters. Although the Classification Table reveals improvement in the *Cauchit* Model, there are indications suggesting that for the model could perform better if the middle category of food consumption group were collapsed into the first category and the model fitted as a binary logistic regression model. Overall, there is sufficient evidence to fit a final Ordinal Regression model for predicting the ordered values of food consumption groups, namely all those in Table 5.10 with significant *p*-values.

5.3 Fitting of Linear Regression model to the continuous response variable

The Linear Regression model, also known as Ordinary Linear Regression, is fitted to a dataset with dependent variable measured on a continuous scale and one or more independent variables. Estimates of coefficients of linear equation arising from fitting the model are given and results of model fit tests and correlation matrices are given to determine the strength of relationships and predictors.

5.3.1 Important assumptions of the Linear Regression model

There are three important assumptions that the Linear Regression is based on. The first and fundamental assumption of the Linear Regression is that there is a straight line relationship between the dependent variable and each predictor. For example, for an independent random variable X_i, a linear relationship with a dependent variable Y_i takes the form of the equation

$$y_i = b_0 + b_1 x_{i1} + \cdots + b_p x_{ip} + \varepsilon_i$$

where y_i is the value of the i^{th} case of the dependent scale variable; p is the number of predictors; b_j is the value of the j^{th} coefficient, $j=0,\ldots,p$; x_{ij} is the value of the i^{th} case of the j^{th} predictor and ε_i is the error in the observed value for the i^{th} case.

The assumption of linearity dictates that a one unit increase in the value of the j^{th} predictor results in an increase in the value of the dependent variable by b_j units. Note that the term b_0 is also called the *intercept* and represents a constant value of the dependent variable when the values of the predictors are equal to 0.

The second assumption is that the *error term* ε_i has a normal distribution with zero mean and that this distribution has a constant variance (or $\sigma^2 = 1$) otherwise the model suffers from a aspect known as *heteroscedasticity*, implying lack of constant of variance (Hosmer and Lemeshow, 2000). The third and last assumption is that the error term value for each case (household or individual) is independent of values of the model variables in the model as well as the values of the error term for other cases (SPSS, 2006).

5.3.2 Exploration of linear relationship

An often used measure for examining a linear relationship between scale variable to examine whether Linear Regression is a suitable model, is by means of a scatter plot. Two scatter plots have been produced to examine the possibility of linear relationship with household size and Wealth Index Score as depicted in Figure 5.10 below.

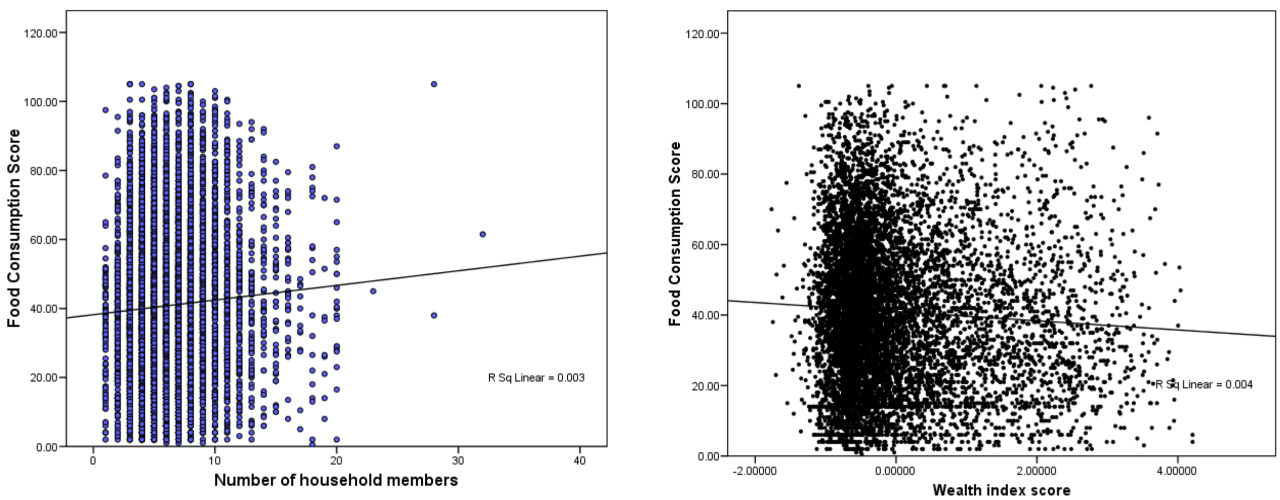

Figure 5.10: Scatter Plots of Food Consumption Scores (*FCS*) by Household Size and *FCS* by Wealth Index Score

From the two plots alone, it is imperative to get an impression that the dataset is not suitable for use of Linear Regression Model. There is too much variability. Nevertheless, the performance of the fitted linear model can be further explored from the SPSS output.

5.3.3 Inspection of the fitness of the model

Table 5.14 displays an edited SPSS output of Analysis of Variance (ANOVA) results. It is shown that the *F* Statistic is significant indicating a good fit model. However, the model summary information showing an R-Square value of 0.179 indicates that the model explains only about 18 *per cent* of the variation. Hence, there is lack of good-fit.

Table 5.14 ANOVA table

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	240383.219	18	13354.623	34.183	.000
	Residual	1100534.452	2817	390.676		
	Total	1340917.671	2835			

5.3.4 Interpretation of the model coefficients

Applying the *backward elimination* selection method described in Chapter 4, the SPSS Regression procedure eliminates non-significant predictors sequentially in seven steps starting with the variable having the highest *p*-value. Hence *sex of household head* was removed in the first step, followed by *experience of food shock*, and ending with *main source of sorghum and millet* with a significance value of 0.079. Recall that almost the same variables were also found to be non-significant, except for *land planted previous year*, by the Ordinal Logistic Regression procedure *PLUM* using the *Complementary Log-Log* link function.

Table 5.15 Part of SPSS output of Linear Regression estimates of coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	52.982	5.752		9.210	.000
	State	-.261	.057	-.091	-4.565	.000
	Type of Household	.677	.612	.019	1.106	.269
	Number of household members	.359	.133	.047	2.705	.007
	Level of education attained	-.423	.391	-.019	-1.082	.279
	Sex of household head	-.056	1.103	-.001	-.050	.960
	Owned land for agriculture	6.250	1.044	.114	5.987	.000
	Used land for agriculture	10.460	1.482	.134	7.057	.000
	Land planted previous year	-.522	.859	-.011	-.607	.544
	Owned livestock	-6.520	.804	-.147	-8.106	.000
	Usually migrated	-5.286	.863	-.111	-6.126	.000
	No. of harvests in one year	3.960	.967	.076	4.095	.000
	Months food last	1.178	.160	.133	7.374	.000
	Vegetable garden	-5.780	.794	-.132	-7.283	.000
	Main sources of livelihood	-.956	.142	-.119	-6.726	.000
	Meals per day (adults and Children)	3.466	.565	.111	6.136	.000
	Main source of sorghum and millet	.490	.278	.032	1.759	.079
	Experienced food shock	.265	.764	.006	.346	.729
	Received food aid	1.045	.841	.023	1.244	.214
	Wealth index quintiles	-.690	.288	-.042	-2.397	.017

5.3.5 Conclusion

There is adequate statistical evidence to suggest that the five independent variables eliminated shown in boxes in Table 5.15 above, do not contribute significantly to the model, whether using the Linear Regression or the Ordinal Regression. However the Linear Regression does worse in explaining variability of fitted values. Again, as explained earlier, the Ordinal Regression is a suitable model to the data because it enables differentiating between the levels of variables in terms of prediction of specific cumulative categories of the response variable food consumption groups.

CHAPTER SIX

CONCLUSIONS AND RECOMMENDATIONS

6.0 Introduction

This chapter attempts to draw general conclusions from the study findings and make specific recommendations on the way forward. The study has generally shown that predicting food insecurity is possible and easy using the appropriate Logistic Regression techniques of common statistics software such as SPSS$^{(R)}$. It is therefore established beyond reasonable doubt that the dataset, which lay idle and unexplored for almost four years, was indeed a missed opportunity. Factors for predicting food insecurity and eliciting timely mitigation efforts were not determined. The study therefore motivates a recommendation that similar statistical modelling procedures should be adopted for analysing data collected using national food security surveys. The study further stimulates a motivation for recommending that factors selected by the statistical modelling technique used, should be adopted in national early warning system interventions for predicting potential food insecurity.

6.1 Conclusions

This project report being a technical document, it is seen appropriate to draw the conclusions based on key findings from the study. The following subsections entail structured conclusions for easy reference.

6.1.1 Proportional Odds Model appropriate for predicting food consumption outcomes

The Proportional Odds Model of the Logistic Regression techniques was found to be appropriate for selecting and determining important factors for predicting the outcome of food consumption and, by extension, food insecurity level based on a similar sample survey. It is further shown that the Logistic Regression model with *Cauchit* link function is more appropriate for producing estimates and significant values used in identification of model predictors. This model can succeed in including and eliminating some factors that the other related models would not; especially the independent variables with marginal significance values. With a carefully selected sample and meticulously supervised data collection, to eliminate or reduce missing cases to a bare minimum, it is guaranteed that the dataset will yield more accurate parameter estimates that can be used in building a model for predicting the risk of food insecurity.

6.1.2 At least eleven factors influenced food insecurity in Southern Sudan

The study in more than one set of analyses determined at least eleven factors as important predictors of food consumption; hence food insecurity in South Sudan (at least during 2006 or immediately after the 21-year civil war ceased). The factors to be included in the model are shown in Appendix 5. Statistical modelling techniques used differed in determination of predictors. The variables and the techniques used for selection are displayed in Table 6.1.

Table 6.1 Variable selection by statistical modelling techniques

	Variable	Cauchit Ordinal Regression Model	Complementary Log-Log Model	Linear Regression Model [a]
1	State	Yes	Yes	Yes
2	Number of household members	**No**	Yes	Yes
3	Months of harvest food lasting	Yes	Yes	Yes
4	Number of meals per day	Yes	Yes	Yes
5	Farmland ownership	Yes[b]	Yes	Yes
6	Farmland use	Yes	Yes	Yes
7	Farmland planting	Yes	Yes	**No**
8	Number of harvests in one year	**No**	Yes	No
9	Livestock ownership	Yes	Yes	Yes
10	Migration/movement of household	Yes	**No**	Yes
11	Ownership of home garden	Yes	Yes	Yes
12	Source of livelihoods	Yes	Yes	Yes
13	Source of sorghum and millet	Yes	Yes	**No**
14	Wealth Index Quintile	**No**	Yes	Yes

[a] Used when the food consumption score is a continuous variable; [b] marginally non-significant

Although the model with *Cauchit* link function is recommended because of presence of extreme values of the food consumption scores, the *Complementary Log-Log* model determines more variables (13) as important predictors. The backward elimination model for fitting a continuous response variable does fairly well too in removing the non-significant (non-influential) variables.

6.1.3 At least eight factors could be used for food insecurity surveillance

By means of Table 6.1 above, it can be concluded that the three techniques are in agreement in determining eight independent variables as important in predicting the outcome of food consumption: *state; number of months in which food harvest lasted; number of meals eaten per day; ownership of farmland; use of farmland; livestock ownership; home garden ownership* and *sources of livelihoods*. With the post-conflict situation of Southern Sudan, some of these factors influenced food consumption negatively. Whereas ownership of certain production assets is supposed to improve the probability of being in better food consumption group, the relationship was quite the opposite. However, this finding provides basis for further investigation.

The eight variables could be included in food insecurity surveillance and routine monitoring exercises applying rapid data collection approaches.

6.1.4 Easily replicable methodology

Quite often agricultural economists tend to ask for quantitative data such as number of acres planted, quantity of food harvested, crop yields, quantity of food purchased, quantity of harvest food sold, quantity received from food aid, income earned from sale of crops and amount spent on buying food items. With rampant illiteracy, recall bias and the amount of enumerator training required, attempts to collect quantitative data synonymous to driving the research toward failure. The study shows that even with the heavy and rampant missing data involving quantitative variables, it was possible to fit a model that determined the variables of importance. Indeed such quantitative variable as number of meals eaten per day, household size and number of months of food lasting, are not only easy to recall but they are also important parameters for predicting the outcome of food consumption and household food insecurity. It is, therefore, shown that the methodology is replicable and adaptable for routine surveillance of food insecurity among highly illiterate populations.

6.1.5 Peculiar findings

The study uncovered some peculiar findings which could be typical of Southern Sudan. Of prominent interest is the finding that food aid was not an important determinant of food consumption! In other words, the difference between food aid recipients and non-recipients in their food consumption was not statistically significant. This is a revelation that although a sizeable 35 per cent of the households responded to have received food aid, their food consumption scores did not improve significantly over the non-recipients. This is evidence that reinforces the fact that food received from aid could provide relief from hunger but would not provide the real solution to food insecurity. Similar in peculiarity is the finding that households that reported to have experienced some sort of food shock (41%: 3783) did not differ significantly in their food consumption levels compared to those that did not experience any shock. This is a clear revelation of the endemic nature of food insecurity in Southern Sudan. In other words, poor food consumption and food shocks – at least for the period of the study – were characteristic of Southern Sudan.

Another finding of interest is the significant difference (*p*-value=0.018) between households receiving staple food items in form of “gift” from relatives and “another” source in relation to experiencing poor food consumption. With odds of 6 times getting staples from relatives than of getting from “other” sources being more probable to score low food consumption, it could be concluded that dependence on relatives, although affecting only 1 *per cent* (91 households), is manifestation of severe coping mechanism; hence the need for further investigation. A further cause of bewilderment is that the reported main sources of livelihood being livestock raring, farming and fishing are revealed to increase the probability of being in the lower food consumption categories. This outcome may also need to be further investigated.

6.2 Recommendations

There are three key recommendations of the study stemming from the conclusions. First, there is need to adopt the methodology and modelling techniques used in the study at the national level by statistical agencies as a food insecurity surveillance tool. Secondly, it may be relatively easy, less time consuming and cheaper to conduct rapid monitoring activities for predicting the likelihood of food insecurity occurrence to be based on the twelve independent variables and questions on food groups and food frequency. Indeed this recommendation falls within the folds of the New Partnership for Africa's Development's Framework for African Food Security quests for finding a vigorous measure for its fourth objective: "*Increased quality of diets through diversification of food among the target groups*" (CAADP-FAFS, 2008).

Finally, this study has unearthed adequate evidence for further research. In other words, the study coming immediately after the dawning of peace in the conflict-riddled Sudan has availed indicators readily used as baselines for current and future food security interventions. In particular, the areas outlined in the foregoing section as "peculiar findings" do not only give adequate reason for conducting follow-up studies, but also motivate policy-making around finding alternatives for encouraging rural people to use resources and opportunities at their hands, for improving their livelihoods and food consumption. As it appears as at two years ago, ownership of livestock and land for agriculture, grazing and other forms of livelihoods, was not adequately utilised to improve food consumption of households. In fact, households with agricultural land or livestock were seen more likely to score poorer in food consumption terms.

REFERENCES

ABBADI, K. A. B. & AHMED, A. E. (2006) Brief overview of Sudan economy and future prospects for agricultural development. *Khartoum Food Aid Forum.* Khartoum, Sudan, World Food Programme.

AGRESTI, A. (2002) *Categorical Data Analysis, 2nd ed.,* New York, John WIley and Sons.

AGRESTI, A. (2004) Analysis of Ordinal Categorical Data New York, Wiley.

ALIBER, M. & MODISELLE, S. (2002) Pilot study on methods to monitor household-level food security for the National Department of Agriculture. Pretoria, Human Sciences Research Council.

ASHBY, D., POCOCK, S. J. & SHAPER, A. G. (1986) Ordered polytomous regression: an example relating serum biochemistry and haematology to alcohol consumption. *Appl. Statistics,* 35**,** 289-301.

BANK, W. (1986) World Development Report 1986. Washington DC, The International Bank for Reconstruction and Development.

BENTLY, M. & PELTO, G. (1991) The household production of nutrition. *Social Science & Medicine,* 33**,** 1101-1102.

BICKEL, G., NORD, M., PRICE, C., HAMILTON, W. & COOK, J. (2000) *Guide to measuring household food security,* Alexandria, VA, Office of Analysis, Nutrition, and Evaluation, Food and Nutrition Service, U.S. Department of Agriculture.

BOND, C. (1998) World News Story Page: 1 million people face famine in Sudan, Ethiopia. *CNN Interactive.* Cable News Network, Inc.

BOUDREAU, T. (1998) The food economy approach: a framework for understanding rural livelihoods. *Relief and Rehabilitation Network Paper No. 26.* Overseas Development Institute

CAADP-FAFS (2008) CAADP Pillar III FAFS Score Card. Pietermaritzburg, African Centre for Food Security (ACFS), University of KwaZulu-Natal (UKZN).

CAHN, M. (2002) *Sustainable Livelihoods Approach: Concept and Practice,* Palmerston North, Massey University.

CHOPAK, C. (2000) Early Warning Primer: An Overview of Monitoring and Reporting. *Famine Early Warning Systems Network.* Washington D.C., USAID FEWS Project.

CHRISTIANSEN & BOISVERT (2000) Measuring household food vulnerability: case evidence from Northern Mali. Working Paper 200-05. *Department of Agriculture, Resource, and Managerial Economics.* Cornell University.

CHUNG, K., HADDAD, L., RAMAKRISHNA, J. & RIELY, F. (1997) Alternative approaches to locating the food insecure: qualitative and quantitative evidence from south India. Washington DC, Food Consumption and Nutrition Division, International Food and Policy Institute.

COATES, J., SWINDALE, A. & BILINSKY, P. (2007) Household Food Insecurity Access Scale (HFIAS) for Measurement of Household Food Access: Indicator Guide (V.3). Washington, D.C., Food and Nutrition Technical Assistance (FANTA) Project.

COATES, J., WEBB, P. & HOUSER, R. (2003) Measuring Food Insecurity: Going Beyond Indicators of Income and Anthropometry. Washington, D.C., Food and Nutrition Technical Assistance (FANTA) Project.

COLLET, D. (1991) *Modelling binary data,* London, Chapman & Hall.

COLLET, D. (2003) *Modelling Binary Data,* London, Chapman & Hall/CRC.

COX, D. R. & SNELL, E. J. (1989) *Analysis of Binary Data,* London, Chapman and Hall.

DFID (1999) Sustainable livelihood guidance sheets.

(http://www.livelihoods.org/info/info_guidanceSheets.html.)

DFID (2001) *Sustainable Livelihoods: Guidance Sheets,* London.

FANTA (2003) Project and Food Aid Management (FAM). *Food Access Indicator Review*. Washington, D.C., Food and Nutrition Technical Assistance (FANTA).

FAO-FIVIMS (2002) Measurement and Assessment of Food Deprivation and Undernutrition. Agriculture and Economic Development Analysis Division. *Interciencia.* Rome, FAO.

FAO-SIFSIA (2008) Food Security Update (Jan-March 2008). Juba, Sudan, Sudan Information for Food Security in Action (SIFSIA).

FAO (1996) World Food Summit – Rome Declaration on World Food Security and WFS Plan of Action. Rome, Food and Agricultural Organisation.

FEWS-NET (2007) Southern Sudan Food Security Watch, February 9, 2007. Nairobi, Kenya, USAID Famine Early Warning Systems Network (FEWSNET).

FIVIMS (2002) Selecting Indicators for National FIVIMS. Rome, Food and Agricultural Organisation.

FRANKENBERGER, T. (1992) Indicators and data collection methods for assessing household food security. IN MAXWELL, S. & FRANKENBERGER, T. (Eds.) *Household food security: concepts, indicators, measurements: A technical review*. New York and Rome, UNICEF and IFAD.

FRONGILLO, E. & NANAMA, S. (2006) Development and validation of an experience-based tool to directly measure household food insecurity within and across seasons in northern Burkina Faso. *The Journal of Nutrition*, 136**,** 1409-1419.

FSAU (2006) Integrated Food Security and Humanitarian Phase Classification: Technical Manual Version I. *Technical Series Report.* Nairobi, Kenya, Food Security Analysis Unit (FSAU).

GIBSON, R. S. (1990) *Principles of nutritional assessment,* Oxford, UK, Oxford University Press.

GILLESPIE, S., HADDAD, L. & JACKSON, R. (2001) HIV/AIDS, food and nutrition security: impacts and actions. *28th Session of the ACC/SCN Symposium on Nutrition and HIV/AIDS.* Washington DC and Rome, International Food and Policy Research Institute and World Food Programme.

GITTELSOHN, J., MOOKHERJI, S. & PELTO, G. (1998) Operationalizing household food security in rural Nepal. *Food and Nutrition Bulletin,* 19**,** 210-222.

GOSS-MOH (2008) Sudan Household Health Survey Report. Juba, Government of Southern Sudan (GOSS).

GOSS (2007) Sudan Household Health Survey Report. Juba, Government of Southern Sudan (GOSS).

HENDRIKS, S. (2005) The challenges facing empirical estimation of household food (in)security in South Africa, Development Southern Africa. *Development Southern Africa,* 22**,** 103–19.

HOCKING, R. R. (2003) *Methods and applications of linear models: regression and the analysis of variance,* Hoboken, NJ, Wiley.

HODDINOTT, J. (1999a) Choosing outcome indicators of household food security. *Training Materials: food security.* Technical Review No. 7. ed. Washington D.C., International Food and Policy Research Institute (IFPRI).

HODDINOTT, J. (1999b) Operationalizing household food security in development projects: an introduction. *Training materials: food security.* Technical Review No. 1. ed. Washington D.C., International Food Policy Research Institute (IFPRI).

HODDIOTT, J. (1999) Choosing outcome indicators of household food security. *Training Materials: food security.* Technical Review No. 7. ed. Washington D.C., International Food and Policy Research Institute (IFPRI).

HOSMER, D. W. & LEMESHOW, S. (2000) *Applied Logistic Regression,* New York, Wiley.

HRW (1999) Famine in Sudan, 1998. USA, Human Rights Watch.

IGAD (2003) IGAD Strategy. Djibouti, Inter-governmental Authority on Development.

JONES, D. & WHITEHEAD, J. (1979) Sequential forms of the log rank and modified Wilcoxon tests for censored data. *Biometrika,* 66**,** 105-113.

LOEVINSOHN, M. & GILLESPIE, S. (2003) HIV/AIDS, food security and rural livelihoods: understanding and responding. RENEWAL Working Paper No. 2. *RENEWAL Workshop.* The Hague and Washington, D.C., International Service for National Agricultural Research and International Food and Policy Research Institute.

LOKOSANG, L. B. (2009) Household Resilience and Community Resilience Indexes for rapid and predictive measurement of household food insecurity risk. Pietermaritzburg, African Centre for Food Security, University of KwaZulu-Natal.

MAXWELL, D., AHIADEKE, C., LEVIN, C., ARMAR-KLEMESU, M., ZAKARIAH, S. & LAMPTEY, G. M. (1999) Alternative food-security indicators: Revisiting the frequency and severity of 'coping strategies'. *Food Policy* 24**,** 411-429.

MAXWELL, D., WATKINS, B., WHEELER, R. & COLLINS, G. (2003) The Coping Strategy Index: a tool for rapidly measuring food security and the impact of food aid programmes in emergencies. *International Workshop on Food Security in Complex Emergencies: building policy frameworks to address longer-term programming challenges.* Rome, Food and Agricultural Organisation (FAO)

MAXWELL, D. G. (1995) Measuring food insecurity: the frequency and severity of ‘coping strategies’. *Food Consumption and Nutrition Division Paper No. 8.* Washington D.C., International Food Policy and Research Institute (IFPRI).

MAXWELL, S. & SMITH, M. (1992) *Household Food Security: A Conceptual Review,* Brighton, UK., Mimeo.

MCCULLAGH, P. (1980) Regression Models for Ordinal Data. *Journal of the Royal Statistical Society B,* 42**,** 109-142.

MCCULLAGH, P. & NELDER, J. A. (1989) *Generalized Linear Models*, London, Chapman & Hall.

MENARD, S. (1995) Applied Logistic Regression Analysis. *Sage University Paper series on Quantitative Applications in the Social Sciences.* Thousand Oaks, CA, Sage.

MENARD, S. (Ed.) (2002) *Applied Logistic Regression Regression Analysis. 2nd edition,* Thousand Oaks, CA, Sage University Press.

MORRIS, S. (1999) Measuring nutritional dimensions of household food security. *Technical Guide No. 5.* Washington D.C., International Food Policy Research Institute.

NANDY, S., KELLY, M., GORDON, D., SUBRAMANIAN, S. & SMITH, G. (2003) Food & Health in Welfare: Current Issues, Future Perspectives. Bristol, UK, Townsend Centre for International Poverty Research, University of Bristol.

PETERSON, B. & HARREL, F. (1990) Partial proportional odds models for ordinal response variables. *Applied Statistics,* 39**,** 205-217.

ReSAKSS (2008) ReSAKSS annual trends report 2008: Monitoring agricultural sector performance, growth, and poverty in Africa. *Annual Trends Report 2008.* Washington, DC., Regional Strategic Analysis and Knowledge Support Systems (ReSAKSS), New Partners for Africa Development (NEPAD).

RIELY, F., MOCK, N., COGIL, B., BAILEY, L. & KENEFICK, E. (1999) Food security indicators and framework for use in the monitoring and evaluation of food aid programs. *Food and Nutrition Technical Assistance (FANTA).* Washington, D.C., Food and Nutrition Technical Assistance Project (FANTA).

ROSEGRANT, M. W. & CLINE, N. A. (2003) Global Food Security: Challenges and Policies. *Science,* 302**,** 1916-1919.

SAS (2008) *SAS Guide for Personal Computers, Version 9.2,* Cary, NC, SAS Institute Inc. (www.support.sas.com/documentation/).

SCF(UK) (2000) *Household Economy Approach: A resource manual for practitioners,* London, Save the Children Fund (SCF) United Kingdom.

SCHLESSELMAN, J. J. (1982) *Case-Conrol Studies: Design, Conduct, Analysis.,* New York, Oxford University Press.

SETBOONSARNG, S. (2005) Child Malnutrition as a Poverty Indicator: An Evaluation in the Context of Different Development Interventions in Indonesia. Asia Development Bank Institute (ADBI) Discussion Paper No. 21.

SINGH, S. (2004) *Market orientation, corporate culture and business performance,* Aldershot, Ashgate.

SMART (2006) Measuring mortality, nutritional status and food security in crisis situations: SMART Methodology.

SPSS (2006) SPSS Version 15.0 for Windows. Chicago, SPSS Inc.

STATSOFT (2007) Electronic Statistics Textbook., Tulsa, OK: http://www.statsoft.com/textbook/stathome.html.

SWINDALE, A. & BILINSKY, P. (2006) Household Dietary Diversity Score (HDDS) for measurement of Household Food Access: Indicator Guide. Washington D.C., Food and Nutrition Technical Assistance (FANTA) Project.

SWINDALE, A. & OHRI-VACHASPATI, P. (2005) Measuring household food access: A Technical Guide. Washington D.C., Food and Nutrition Technical Assistance (FANTA) Project.

UN (2000) United Nations Millennium Declaration. *8th Plenary Meeting.* New York, United Nations General Assembly.

UNDP (2009) *Human Development Report 2009: Overcoming barriers,* New York, NY, Palgrave Macmillan.

UNICEF (2008) Multiple Indicator Cluster Surveys (MICS). New York, NY, Statistiscs and Monitoring Division of Policy and Practices, UNICEF (www.childinfo.org/mics.html).

WEBB, P. & ROGERS, B. (2003) Addressing the “In” in Food Insecurity. Occasional Paper No. 1. Washington D.C., USAID Office of Food for Peace.

WEBB, P. J., COATES, E., FRONGILLO, B., ROGERS, A., SWINDALE, A. & BILINSKY, P. (2006) Measuring Household Food Insecurity: Why It’s So Important and Yet So Difficult to Do. Washington D.C., Food and Nutrition Technical Assistance (FANTA) Project.

WFP-CFSVA (2007) CFSVA Methodology Workshop Report. Rome, WFP.

WFP-VAM (2007a) Sudan: Comprehensive food security and vulnerability analysis (CFSVA). Rome, World Food Programme.

WFP-VAM (2007b) Sudan: Southern Sudan Comprehensive Food Security and Vulnerability Analysis (CFSVA). Rome, Italy, World Food Programme.

WFP-VAM (2008) Food consumption analysis: Calculation and use of the food consumption score in food security

analysis. Rome, World Food Programme, Vulnerability Analysis and Mapping Branch (ODAV).

WHO (1995) Physical status: the use and interpretation of anthropometry. Geneva, WHO Technical Report No. 854.

WOLFE, S. W. & FRONGILLO, E. A. (2000) Building household food security measurement the ground up: background paper. Washington D.C., Food and Nutrition Technical Assistance Project.

WOOLFE, B. (1955) On estimating he relation between blood group and disease. *Ann. Human Genetics*, 19**,** 251-3.

WORDNET (2006) El nino. *WordNet.* Princeton, http://wordnet.princeton.edu.

APPENDIX 1

FOOD SECURITY QUESTIONNAIRE USED IN THE DATA COLLECTION

SUDAN HOUSEHOLD HEALTH SURVEY

food security questionnaire

I would now like to ask you questions about your household, Livelihood and Food Security

This module is to be administered to the household head or spouse of the household head (same person interviewed for the household questionnaire). Please make sure to complete the household number on top of pages.

HOUSEHOLD CIRCUMSTANCES		**HCI**
HCI 1. WHICH OF THE FOLLOWING BEST DESCRIBES YOUR HOUSEHOLD? ***(READ ANSWERS, CIRCLE ONLY ONE)***	Internally Displaced .... 1 Refugee .... 2 Returnee ex-Internally Displaced ..... 3 Returnee ex Refugee .... 4 Resident 5	
HCI 2. DID YOUR HOUSEHOLD LIVE HERE **12 MONTHS AGO** (1 YEAR)?	Yes .... 1 No 2	**1⇨Skip to HCI 5**
HCI 3. WHERE DID YOUR FAMILY LIVE BEFORE YOU MOVED TO THIS LOCATION?	Nearby Village (<10km).................... 1 Distant Village (>10km) .... 2 IDP Camp .... 3 Other State .... 4 Other Country 5	
HCI 4. HOW MANY MONTHS AGO DID YOU MOVE TO THIS CURRENT LOCATION ***(WRITE 98 IF DON'T KNOW)***	____ ____ months	
HCI 5. DOES YOUR HOUSEHOLD USUALLY MIGRATE DURING THE YEAR FOR WORK OR TO RAISE LIVESTOCK?	Yes .... 1 No 2	

HOUSEHOLD BELONGINGS AND LIVESTOCK		HBL
HBL 1. Does your household own any of the following belongings (assets) in working	Chair .... A Table .... B Bed C	

condition? *(CIRCLE ALL THAT APPLY)*	Lantern .... D Cooking Utensils .... E Hoe .... F Axe .... G Ox Drawn Plough .... H Hand Hammer Mill .... I Hammer Mill L	
HBL 2. DOES YOUR HOUSEHOLD OWN ANY LIVESTOCK, HERDS OR FARM ANIMALS (EVEN IF THEY ARE NOT THERE NOW)? *COPY ANSWER FROM HI3, HOUSEHOLD MODULE*	Yes........ No	**2⇨Skip to HBL 10**
HBL 3. How many **cattle** does this household have? *COPY ANSWER FROM HI4, HOUSEHOLD MODULE*	0 .... 1 – 5........ 6 – 20........ 21 – 50 .... 51 – 100 .... More than 100........ Don't Know	
HBL 4. How many **milk cows** does this household have? *COPY ANSWER FROM HI7, HOUSEHOLD MODULE*	0 .... 1 – 4........ 5 – 9........ 10 – 14 .... 15 – 20 .... More than 20........ Don't Know	
HBL 5. How many **chickens** does this household have? *COPY ANSWER FROM HI5, HOUSEHOLD MODULE*	0 .... 1 – 10........ 11 – 20........ 21 – 50 .... 51 – 100 .... More than 100........ Don't Know	
HBL 6. How many **goats** does this household have? *COPY ANSWER FROM HI6, HOUSEHOLD MODULE*	0 .... 1 – 5........ 6 – 20........ 21 – 50 .... 51 – 100 .... More than 100........ Don't Know	
HBL 7. How many **sheep** does this household	0 .... 1 – 5........ 6 – 20........ 21 – 50	

have? *COPY ANSWER FROM HI8, HOUSEHOLD MODULE*	51 – 100 .... More than 100.......... Don't Know	
HBL 8. How many **horses, donkeys or mules** does this household have? *COPY ANSWER FROM HI9, HOUSEHOLD MODULE*	0 .... 1 – 3.......... More than 3.......... Don't know	
HBL 9. How many **camels** does this household have? *COPY ANSWER FROM HI10, HOUSEHOLD MODULE*	0 .... 1 – 3.......... More than 3.......... Don't know	
HBL 10. Did your household sell any belongings or livestock in **the last 1 year**?	Yes.......... No	2⇨Skip to **Module LAP**
HBL 11. Which was the main reasons for selling belongings and livestock? ***(CIRCLE ALL THAT APPLY)***	To Eat / Purchase Food.......... To Pay medical Expenses.......... To Repay Debts.......... For Social Events .... To Pay Normal Daily Expenses .... For School Fees / Expenses .... To Purchase Agricultural Inputs .... Other, specify _______________..........	

LIVELIHOODS AND AGRICULTURAL PRODUCTION		LAP
LAP 1. DOES YOUR HOUSEHOLD USUALLY USE LAND FOR FARMING?	Yes..........1 No 2	**2⇨Skip to LAP 6**
LAP 2. HOW MUCH LAND DO YOU HAVE ACCESS TO FOR FARMING? *(write "98" for don't know)*	___ ___ feddans (**1 feddan = 0.42 ha**)	
LAP 3. HOW MANY HARVESTS CAN YOU TYPICALLY HAVE IN ONE YEAR?	1 2	
LAP 4. HOW MANY MONTHS DOES FOOD FROM YOUR HARVEST TYPICALLY LAST?	___ ___ months	
LAP 5. HOW LONG DOES THE HUNGER SEASON TYPICALLY LAST?	___ ___ months	

LAP 6. DID YOU PLANT LAND IN THE **PREVIOUS SEASON**?	Yes..1 No 2	**2⇨Skip to LAP 12**
LAP 7. HOW MUCH LAND DID YOU PLANT IN THE **PREVIOUS SEASON**?	___ ___ feddans (**1 feddan = 0.42 ha**)	

Please after completing LAP 1-7, fill in the following table 1 crop at a time (line by line), repeating the top questions

LAP 8. WHAT ARE THE MAIN CROPS CULTIVATED BY YOUR HOUSEHOLD IN THE PREVIOUS SEASON? (***USE CROP CODE BELOW, WRITE 88 IF NONE***)	LAP 9. HOW MUCH DID YOU PRODUCE IN THE PREVIOUS SEASON? **UNITS CODE** 1 = BAG (90 KG) 2 = SACK (50 KG) 3 = TIN (15 KG) 4 = MALUA (3 KG) 5 = PIECE	LAP 10. WHAT IS THE PROPORTION OF THIS CROP THAT YOU CONSUME IN THE PREVIOUS SEASON? (**USE PROPORTIONAL PILLING METHOD**)	LAP 11. WHAT IS THE PROPORTION OF THIS CROP THAT YOU SOLD AND / OR EXCHANGED (BARTER) IN THE PREVIOUS SEASON? (**USE PROPORTIONAL PILLING METHOD**)
a. Main Crop: ___	__ __ __ units: __	___ ___ ___ %	___ ___ ___ %
b. Second Crop: ___	__ __ __ units: __	___ ___ ___ %	___ ___ ___ %
c. Third Crop: ___	__ __ __ units: __	___ ___ ___ %	___ ___ ___ %
d. Fourth Crop: ___	__ __ __ units: __	___ ___ ___ %	___ ___ ___ %
e. Fifth crop: ___	__ __ __ units: __	___ ___ ___ %	___ ___ ___ %

CROP CODE:
1. SORGHUM
2. MILLET
3. MAIZE
4. RICE
5. OTHER CEREALS
6. CASSAVA
7. SWEET POTATOES
8. BEANS
9. COWPEAS
10. SESAME
11. GROUNDNUTS
12. PUMPKIN
13. OTHER VEGETABLES
14. WATERMELON
15. OTHER FRUIT
14. TOBACCO
15. OTHER CASH CROP
16. OTHER CROP (SPECIFY) ___________

LAP 12. DOES YOUR HOUSEHOLD HAVE A VEGETABLE PLOT / GARDEN?	Yes .... 1 No.. 2	

LAP 13. WHICH WERE THE MAIN SOURCES OF LIVELIHOOD IN THE LAST **1 YEAR**? (***USE INCOME SOURCES CODE BELOW; WRITE 88, IF NONE***)

a. Main Source of livelihood: ___ ___	**INCOME SOURCE CODE:**
b. Second Source of livelihood: ___ ___	
c. Third Source of livelihood: ___ ___	
d. Fourth Source of livelihood: ___ ___	
e. Fifth Source of livelihood: ___ ___	

INCOME SOURCE CODE:
1. LIVESTOCK REARING
2. AGRICULTURAL PRODUCTION
3. FISHING
4. HUNTING, GATHERING
5. PETTY TRADE (BREWING / SMALL BUSINESS)
6. COLLECTION OF NATURAL RESOURCES (FIREWOOD, GRASS, ...)
7. UNSKILLED LABOR (FARM JOB, …)
8. HANDICRAFT (BASKETS ,…)
9. SKILLED LABOR (ARTISAN)
10. EMPLOYEE WORK (SALARIED, TEACHER,...)
11. FOOD AID ASSISTANCES
12. OTHER ACTIVITY? SPECIFY ___________

LAP 14. DO YOU USUALLY WORK FOR FOOD ONLY (NO PAYMENT)?	Yes..1 No 2

HOUSEHOLD EXPENDITURES — HEX

Complete one bought item at a time (line by line)			
ITEM	HEX 1. IN THE LAST **<u>3 MONTHS</u>**, DID YOU BUY (*ITEM*)?	HEX 2. DID YOU PAY CASH FOR IT OR PART OF IT?	HEX 3. DID YOU USE BARTER / EXCHANGE FOR PART OF IT OR FOR THE TOTAL?
Cereals (Sorghum, maize, millet,…)	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Roots and Tubers (Cassava, groundnut, …)	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Pulses, vegetables and fruits (Beans, pumpkin, mango…)	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Meat and fish	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Sugar, salt and cooking oils	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Cooking fuel, lighting	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Alcohol and Tobacco	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Grinding, Milling	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Medical services and items (drugs,…)	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Education, school fees and school materials	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Clothing, shoes	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Equipment, tools, seeds	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Hiring labor (farm hand, construction,..)	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
House construction and repair materials	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No
Fines, taxes, debts and rents	1. Yes 2. No ⇨**Next item**	1. Yes 2. No	1. Yes 2. No

FOOD CONSUMPTION AND SOURCES FCS		
FCS 1. In a <u>normal/usual</u> period, how many times do the <u>adults</u> (over 15) in this household eat in a day?	0 1 2 3 4 5	
FCS 2. In a <u>normal/usual</u> period, how many times do the <u>children</u> (15 or under) in this household eat in a day?	0 1 2 3 4 5	
FCS 3. In a <u>hunger</u> period, how many times do the <u>adults</u> (over 15) in this household eat in a day?	0 1 2 3 4 5	
FCS 4. In a <u>hunger</u> period, how many times do the <u>children</u> (15 or under) in this household eat in a day?	0 1 2 3 4 5	

Complete one item consumed at a time (line by line), each time repeating the question

Food Item	FCS 5. Last week, how many **days** did your household eat this item?	FCS 6. Compared to last week, how much did you eat this item in the last <u>harvest season</u>?	FCS 7. Compared to last week, how much did you eat this item in the last <u>rainy season</u>?	FCS 8. What was the main source for this item? **<u>Food Sources Codes:</u>** 1 = Own Production 2 = Market Purchase 3 = Hunting, Fishing, Gathering 4 = Exchange (labor / items) 5 = Borrowed 6 = Gift (family, relatives) 7 = Food Aid 8 = Other 9 = do not know
a. Sorghum and millet	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
b. Maize	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
c. Roots and tubers (cassava, yams, potatoes)	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
d. Pulses (beans, peas, lentils, ...)	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
E. Okra	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
f. Groundnuts	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
g. Sesame	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
h. Wild plants and vegetables (leaves, fruits, greens...)	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
i. Meat	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___

J. FISH	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
K. EGGS	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
L. MILK	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
M. OIL, FAT, BUTTER	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___
N. SUGAR	0 1 2 3 4 5 6 7	1 - 2 - 3 Same -More- Less	1 - 2 - 3 Same -More- Less	Code: ___

SHOCKS AND COPING MECHANISMS — SCM

SCM 1. DURING THE LAST **1 YEAR**, DID YOUR HOUSEHOLD EXPERIENCE ANY INCIDENT THAT AFFECTS ITS USUAL ABILITY TO EAT AND/OR BUY FOODS OF THE QUALITY, QUANTITY OR VARIETY YOU PREFER?	Yes…………………………1 No………………………….. 2	**2⇨Skip to Module FAI**

After completing question SCM1, complete one incident at a time (line by line), each time repeating the questions above

SCM 2. BY ORDER OF IMPORTANCE, WHAT INCIDENTS DID YOUR HOUSEHOLD EXPERIENCE IN THE LAST **1 YEAR**	SCM 3. WHAT IS THE MAIN ACTION YOUR HOUSEHOLD TOOK TO COMPENSATE THE EFFECT OF THAT INCIDENT?	SCM 4. HOW OFTEN DID YOU DO THIS IN THE LAST **1 YEAR**?	SCM 5. DID YOUR HOUSEHOLD RECOVER FROM THAT INCIDENT?
INCIDENTS CODE: 1 = INSECURITY, VIOLENCE 2 = INCREASED PRICE FOR FOOD 4 = DROP IN FARM GATE PRICE 5 = FLOODS 6 = DROUGHT/DRY SPELL 7 = CROP PEST AND DISEASE 8 = LIVESTOCK DISEASE 9 = SICKNESS OF HOUSEHOLD MEMBER 10 = DEATH OF HOUSEHOLD MEMBER 11 = INCREASED HOUSEHOLD SIZE (IDPS) 12 = LOSS / LACK OF EMPLOYMENT	**COPING CODE:** 0 = NOTHING 1 = EAT LESS PREFERRED FOODS 2 = EAT FEWER OR SMALLER MEALS PER DAY 3 = GO ONE ENTIRE DAY WITHOUT MEALS 4 = COLLECT WILD FOODS, HUNT OR HARVEST IMMATURE CROPS 5 = DISTRESS SALE / SLAUGHTER OF LIVESTOCK 6 = DISTRESS SALE OF OTHER ASSETS 7 = PURCHASE FOOD ON CREDIT 8 = BORROW FOOD FROM FAMILIES AND FRIENDS, KINSHIP SUPPORT 9 = WORKED FOR MONEY 10 = WORKED FOR FOOD ONLY 11 = REDUCED EXPENDITURES ON HEALTH OR EDUCATION 12 = SPENT SAVINGS 13 = SOME HOUSEHOLD MEMBERS MIGRATED 14 = OTHER (SPECIFY)______________		
MAIN : (Code) ___	(Code) ___	___ times	1. Yes 2. No
SECOND : (Code) ___	(Code) ___	___ times	1. Yes 2. No
THIRD : (Code) ___	(Code) ___	___ times	1. Yes 2. No
FOURTH : (Code) ___	(Code) ___	___ times	1. Yes 2. No
FIFTH : (Code) ___	(Code) ___	___ times	1. Yes 2. No

FAI 1. DID YOUR HOUSEHOLD RECEIVE FOOD AID IN THE	Yes	**2⇨Skip to FAI 10**

LAST **3 MONTHS**? (SINCE BEGINNING OF THE YEAR)	No	
FAI 2. HOW MANY TIMES DID YOU RECEIVE FOOD AID IN THE LAST **3 MONTHS** (SINCE BEGINNING OF THE YEAR)?	____ times	
FAI 3. WHAT FOODS DID YOU RECEIVE DURING THE LAST **3 MONTHS**? (SINCE BEGINNING OF THE YEAR) ***(CIRCLE ALL THAT APPLY)***	Cereals .... Pulses..................................... CSB .... Vegetable Oil .... Sugar .... Salt	
FAI 4. DID YOU TRADE OR SELL ANY OF THE COMMODITIES YOU RECEIVED AS FOOD AID?	Yes .... No	**2⇨Skip to FAI 6**

FAI 5. WHY DID YOU TRADE OR SELL THE COMMODITIES YOU RECEIVED AS FOOD AID? ***(CIRCLE ALL THAT APPLY)***	To buy/get other foods..................... A To purchase animals .... B To pay agricultural expenses........... C To pay for milling............................. D To pay for education expenses........ E To pay for health expenses .... F To purchase fuel .... G To repay debts.................................. H Other, specify________________.......I	
FAI 6. IN GENERAL, HOW MANY WEEKS DOES FOOD AID LAST AFTER IT IS DISTRIBUTED? ***(IF LESS THAN 1 WEEK, WRITE 0; IF DON'T KNOW, WRITE 98)***	__________ weeks	
FAI 7. WHO COLLECTS FOOD AID? ***(CIRCLE ALL THAT APPLY)***	Male .... A Female .... B Children.. C	
FAI 8. WHO IN THE HOUSEHOLD MAKES DECISION ABOUT THE USE OF FOOD AID COMMODITIES?	Male ....1 Female ....2 Both male and female.......................3	
FAI 9. HOW LONG DOES IT TAKE YOU TO GET FROM YOUR HOUSE TO THE FOOD DISTRIBUTION POINT-ONE WAY? ***(IF LESS THAN 1 HOUR, WRITE 0; IF DON'T KNOW, WRITE 98)***	____ walking hours	
FAI 10. DID YOU RECEIVE SEEDS IN THE LAST AGRICULTURAL/FARMING SEASON?	Yes..1 No 2	
FAI 11. DID YOU RECEIVE HAND TOOLS AND/OR PLOUGH IN THE LAST AGRICULTURAL/FARMING SEASON?	Yes..1 No 2	

INSTRUCTIONS FOR ENUMERATOR

SI3. Does any eligible woman aged 15-49 reside in the household? Check Household questionnaire - household listing, column HL6.You should have a questionnaire with the Information Panel filled in for each eligible woman. ☐ Yes. ⇨ Go to QUESTIONNAIRE FOR INDIVIDUAL WOMEN to administer the questionnaire to the first eligible woman. ☐ No. ⇨ Continue.
SI4. Does any child under the age of 5 reside in the household? Check Household questionnaire - household listing, column HL7. You should have a questionnaire with the Information Panel filled in for each eligible child. ☐ Yes. ⇨ Go to QUESTIONNAIRE FOR CHILDREN UNDER FIVE to administer the questionnaire to caretaker of the first eligible child. ☐ No. ⇨ End the interview by thanking the respondent for his/her cooperation. Gather together all questionnaires for this household and tally the number of interviews completed on the cover page of the Household Questionnaire.

APPENDIX 2

LIST OF INDEPENDENT VARIABLES INCLUDED IN THE FIRST MODEL

	Variable description	Code*	Reason for Inclusion
1	State	HH1	States are units of monitoring and focus of economic policy making. Sharp disparities in food consumption could arouse concerns and instigate further investigation and action by relevant federal and state authorities.
2	Household type	HCI1	Households might differ in their consumption levels according to whether they are displacement or not
3	Number of household members	HH11	It is possible that food consumption is strained by bloated household sizes. It is also possible that a big household size could present an opportunity in that there could be more working adults and hence more sources of food. Either way, food consumption might be affected.
4	Sex of household head	HL4	Are male-headed household better off in food consumption levels that female-headed, or vice-versa?
5	Level of education of household head	ED3	Are household headed by a person with higher education level significantly better off than those of lower education?
6	Ownership of land agricultural purposes (farming, grazing or fishing)	HI1	It is possible that land tenure presented an opportunity for households to have better income and food sources than those without. It is also possible that such reasons are not tapped and hence there is no significant difference between owning a piece of land and not owning one.
7	Use of land for farming	LAP1	Households using land for farming are likely to produce food and can turn farm produce to income and improve access to food than households with no land or not using it. FCS might differ significantly between these types of households.
8	Land planted previous season	LAP6	Households where land was planted in the previous season might have enough stocks and have better FCS score than those who did not. It is good to determine whether this is the case or not.

	Variable description	Code*	Reason for Inclusion
9	Ownership of livestock	HBL2	Livestock ownership could be a ready source of food high in micronutrients and protein and thus improving the dietary diversity and food consumption frequency of households over those that do not own livestock.
10	Usual migration of households	HCI5	Instability of households could cause strains on the household budget and food consumption.
11	Number of harvests in one year	LAP3	Farming households harvesting twice in a year, should have food available throughout and could have a better dietary diversity of micronutrient rich food than others. Could there be a significant difference between harvesting twice and harvesting once?
12	Months harvest food lasted	LAP4	It is worthwhile finding out how households fared in terms of food consumption with regard to stocking farmed food.
13	Availability of vegetable plot or home garden	LAP12	Availability of a plot of land or a vegetable garden increases the potential of the household's dietary intake of vitamin rich foods. Thus the FCS is expected to be better for those families than those without home gardens.
14	Main sources of livelihood	LAP13	It is out of question that main source of livelihood is supposed to influence food consumption positively in that the stronger the weight of the main source, the better the access to diversity of diets.
15	Number of meals per day	FCS1/2	Having more meals a day increases the diversity of diets. Hence, it is probably a matter of knowing the magnitude of the influence and significance of this variable to food consumption levels.
16	Main source of sorghum and millet	FCS8	Cereals represent the main staple for most countries. It is therefore important to investigate which source(s) improve or worsen the probability of being in a better food consumption group.
17	Experience of shock or strain	SCM1	It is expected that households that have experienced some shock or a strain might accordingly experience inadequate dietary intake. This possible disparity between the two with regards to the effect on FCS needs to be investigated.
18	Incidence of receipt of food aid in the last 3 months	FAI1	It is worthwhile investigating whether receipt of food aid had a significant effect on FCS. It is normal to expect that a household receiving food aid has better FCS than one that does not. This notion could be wrong in some circumstances when the household entirely dependent on food aid has lower dietary diversity than a household that does not receive food aid. The latter might own home gardens and other more preferred means of livelihoods.
19	Wealth index quintiles	None	Wealth Index Quintiles are proxy measures of poverty and hence, by extension, indirect measures of food insecurity. Wealth index Quintiles are calculated out of a range of household assets owned. It is worth determining the influence of this variable over the household's food consumption levels and finding out which levels of the index have statistically significant influence.

* Code as in the survey questionnaire.

APPENDIX 3

SPSS CODE FOR ANALYSIS OF MODELING OF THE FOOD SECURITY DATA

```
*/ METHOD 1: The Ordinal Regression technique using the Complementary Log-Log
Link Function. The dependent variable is ordinal categorical.

PLUM

  FCG  BY State HHType Edlevel SexHHH OwnLand UseLand PlantLand LStock

  Migrates Harvests VGarden LivSource CerealSrce FoodShk FoodAid WIQntile

   WITH HHsize FoodLast Meals

  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5) PCONVERGE

 (1.0E-6) SINGULAR(1.0E-8)

  /LINK = CLOGLOG

  /PRINT = FIT PARAMETER SUMMARY TPARALLEL

  /SAVE = PREDCAT .

*/ METHOD 2: The Ordinal Regression technique using the Complementary Cauchit
Link Function. The dependent variable is ordinal categorical.

GET FILE='E:\UKZN_MSc\DataSetFinal.sav'.

DATASET NAME DataSet1 WINDOW=FRONT.

PLUM

  FCG  BY State HHType Edlevel SexHHH OwnLand UseLand PlantLand LStock

  Migrates Harvests VGarden LivSource CerealSrce FoodShk FoodAid WIQntile

   WITH HHsize FoodLast Meals

  /CRITERIA = CIN(95) DELTA(0) LCONVERGE(0) MXITER(100) MXSTEP(5) PCONVERGE

 (1.0E-6) SINGULAR(1.0E-8)

  /LINK = CAUCHIT

  /PRINT = FIT PARAMETER SUMMARY TPARALLEL
```

```
  /SAVE = PREDCAT .
```

*/ METHOD 3: The Linear Regression technique for fitting a model where the dependent variable is a ratio scale (continuous). Note that the BACKWARD ELIMINATION CRITERIA is used.

```
REGRESSION

  /MISSING LISTWISE

  /STATISTICS COEFF OUTS R ANOVA COLLIN TOL ZPP

  /CRITERIA=PIN(.05) POUT(.10)

  /NOORIGIN

  /DEPENDENT FCS

  /METHOD=BACKWARD State HHType HHsize Edlevel SexHHH OwnLand UseLand

  PlantLand LStock Migrates Harvests FoodLast VGarden LivSource Meals

  CerealSrce FoodShk FoodAid WIQntile  .
```

APPENDIX 4a

SOME EDITED SPSS ORDINAL REGRESSION OUTPUT FOR A MODEL FITTED WITH COMPLEMENTARY LOG-LOG LINK FUNCTION

PLUM - Ordinal Regression

```
[DataSet1] E:\UKZN_MSc\DataSetFinal.sav
```

Warnings

There are 5630 (66.6%) cells (i.e., dependent variable levels by combinations of predictor variable values) with zero frequencies.

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	6025.082			
Final	5413.788	611.294	49	.000

Link function: Complementary Log-log.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	5722.532	5581	.091
Deviance	5411.016	5581	.947

Link function: Complementary Log-log.

Pseudo R-Square

Cox and Snell	.195
Nagelkerke	.221
McFadden	.101

Link function: Complementary Log-log.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Upper Bound	Lower Bound
Threshold	[FCG = 1]	1.648	.524	9.878	1	.002	.620	2.676
	[FCG = 2]	2.528	.525	23.176	1	.000	1.499	3.557
Location	HHsize	.023	.010	5.886	1	.015	.005	.042
	FoodLast	.049	.012	17.248	1	.000	.026	.072
	Meals	.297	.044	46.463	1	.000	.212	.383
	[State=71]	.721	.139	26.998	1	.000	.449	.993
	[State=72]	.739	.136	29.423	1	.000	.472	1.006
	[State=73]	1.123	.145	60.279	1	.000	.840	1.407
	[State=81]	.397	.126	9.986	1	.002	.151	.643
	[State=82]	.399	.122	10.660	1	.001	.159	.638
	[State=83]	.622	.123	25.488	1	.000	.380	.863
	[State=84]	.361	.108	11.194	1	.001	.150	.573
	[State=91]	.236	.107	4.829	1	.028	.025	.446
	[State=92]	1.184	.124	91.397	1	.000	.941	1.426
	[State=93]	0(a)	.	.	0	.	.	.
	[HHType=1]	.142	.100	2.003	1	.157	-.055	.339
	[HHType=2]	.080	.113	.494	1	.482	-.143	.302
	[HHType=3]	0(a)	.	.	0	.	.	.
	[Edlevel=1]	.037	.094	.154	1	.695	-.148	.222
	[Edlevel=2]	.057	.098	.337	1	.562	-.135	.248
	[Edlevel=3]	-.154	.109	1.980	1	.159	-.368	.060
	[Edlevel=4]	0(a)	.	.	0	.	.	.
	[SexHHH=1]	-.029	.078	.139	1	.709	-.182	.124
	[SexHHH=2]	0(a)	.	.	0	.	.	.
	[OwnLand=1]	-.166	.081	4.215	1	.040	-.324	-.008
	[OwnLand=2]	0(a)	.	.	0	.	.	.
	[UseLand=1]	-.381	.118	10.468	1	.001	-.611	-.150
	[UseLand=2]	0(a)	.	.	0	.	.	.
	[PlantLand=1]	.143	.062	5.344	1	.021	.022	.264
	[PlantLand=2]	0(a)	.	.	0	.	.	.
	[LStock=1]	.286	.059	23.310	1	.000	.170	.402
	[LStock=2]	0(a)	.	.	0	.	.	.
	[Migrates=1]	.117	.065	3.274	1	.070	-.010	.244
	[Migrates=2]	0(a)	.	.	0	.	.	.
	[Harvests=1]	-.088	.076	1.346	1	.246	-.236	.060
	[Harvests=2]	0(a)	.	.	0	.	.	.
	[VGarden=1]	.364	.058	40.068	1	.000	.251	.477
	[VGarden=2]	0(a)	.	.	0	.	.	.
	[LivSource=1]	.628	.193	10.642	1	.001	.251	1.006
	[LivSource=2]	.358	.191	3.499	1	.061	-.017	.733
	[LivSource=3]	.479	.237	4.082	1	.043	.014	.943

	[LivSource=4]	-.042	.239	.031	1	.860	-.511	.426
	[LivSource=5]	.167	.246	.460	1	.498	-.316	.650
	[LivSource=6]	-.143	.208	.473	1	.492	-.551	.265
	[LivSource=7]	-.093	.233	.158	1	.691	-.549	.364
	[LivSource=8]	-.129	.303	.183	1	.669	-.722	.464
	[LivSource=9]	.444	.369	1.448	1	.229	-.279	1.168
	[LivSource=10]	.422	.315	1.790	1	.181	-.196	1.040
	[LivSource=11]	-.044	.236	.035	1	.852	-.507	.418
	[LivSource=12]	0(a)	.	.	0	.	.	.
	[CerealSrce=1]	.571	.439	1.692	1	.193	-.289	1.431
	[CerealSrce=2]	.641	.442	2.105	1	.147	-.225	1.507
	[CerealSrce=3]	.725	.506	2.057	1	.152	-.266	1.717
	[CerealSrce=4]	-.192	.513	.140	1	.708	-1.197	.814
	[CerealSrce=5]	-.086	.718	.014	1	.904	-1.492	1.320
	[CerealSrce=6]	1.203	.586	4.217	1	.040	.055	2.350
	[CerealSrce=7]	.551	.450	1.495	1	.222	-.332	1.434
	[CerealSrce=8]	0(a)	.	.	0	.	.	.
	[FoodShk=1]	.077	.055	1.983	1	.159	-.030	.185
	[FoodShk=2]	0(a)	.	.	0	.	.	.
	[FoodAid=1]	-.026	.064	.163	1	.686	-.152	.100
	[FoodAid=2]	0(a)	.	.	0	.	.	.
	[WIQntile=1]	.150	.094	2.546	1	.111	-.034	.335
	[WIQntile=2]	.232	.088	7.014	1	.008	.060	.404
	[WIQntile=3]	.131	.082	2.570	1	.109	-.029	.292
	[WIQntile=4]	.041	.077	.285	1	.593	-.110	.192
	[WIQntile=5]	0(a)	.	.	0	.	.	.

Link function: Complementary Log-log.

a This parameter is set to zero because it is redundant.

Test of Parallel Lines(c)

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	5413.788			
General	5291.444(a)	122.344(b)	49	.000

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a The log-likelihood value cannot be further increased after maximum number of step-halving.

b The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.

c Link function: Complementary Log-log.

APPENDIX 4b

SOME EDITED SPSS ORDINAL REGRESSION OUTPUT FOR A MODEL FITTED CAUCHIT LINK FUNCTION

PLUM - Ordinal Regression

```
[DataSet1] E:\UKZN_MSc\DataSetFinal.sav
```

Warnings

There are 5630 (66.6%) cells (i.e., dependent variable levels by combinations of predictor variable values) with zero frequencies.

Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	6025.082			
Final	5432.723	592.359	49	.000

Link function: Cauchit.

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	5551.447	5581	.608
Deviance	5429.950	5581	.925

Link function: Cauchit.

Pseudo R-Square

Cox and Snell	.189
Nagelkerke	.215
McFadden	.098

Link function: Cauchit.

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Threshold	[FCG = 1]	3.010	.735	16.789	1	.000	1.570	4.450
	[FCG = 2]	4.189	.743	31.788	1	.000	2.733	5.645
Location	HHsize	.016	.012	1.845	1	.174	-.007	.040
	FoodLast	.091	.015	36.377	1	.000	.061	.120
	Meals	.345	.057	37.155	1	.000	.234	.456
	[State=71]	1.007	.184	30.023	1	.000	.647	1.367
	[State=72]	1.137	.181	39.611	1	.000	.783	1.491
	[State=73]	1.482	.190	60.573	1	.000	1.109	1.856
	[State=81]	.570	.170	11.252	1	.001	.237	.903
	[State=82]	.420	.168	6.253	1	.012	.091	.750
	[State=83]	.871	.164	28.099	1	.000	.549	1.193
	[State=84]	.614	.146	17.776	1	.000	.329	.900
	[State=91]	.521	.147	12.649	1	.000	.234	.809
	[State=92]	1.381	.160	74.121	1	.000	1.066	1.695
	[State=93]	0(a)	.	.	0	.	.	.
	[HHType=1]	.074	.129	.331	1	.565	-.178	.327
	[HHType=2]	-.052	.147	.125	1	.724	-.340	.236
	[HHType=3]	0(a)	.	.	0	.	.	.
	[Edlevel=1]	.002	.121	.000	1	.986	-.235	.239
	[Edlevel=2]	.002	.125	.000	1	.984	-.242	.247
	[Edlevel=3]	-.231	.142	2.675	1	.102	-.509	.046
	[Edlevel=4]	0(a)	.	.	0	.	.	.
	[SexHHH=1]	-.041	.100	.169	1	.681	-.237	.155
	[SexHHH=2]	0(a)	.	.	0	.	.	.
	[OwnLand=1]	-.190	.101	3.527	1	.060	-.387	.008
	[OwnLand=2]	0(a)	.	.	0	.	.	.
	[UseLand=1]	-.546	.148	13.607	1	.000	-.837	-.256
	[UseLand=2]	0(a)	.	.	0	.	.	.
	[PlantLand=1]	.218	.080	7.384	1	.007	.061	.375
	[PlantLand=2]	0(a)	.	.	0	.	.	.
	[LStock=1]	.385	.077	24.974	1	.000	.234	.536
	[LStock=2]	0(a)	.	.	0	.	.	.
	[Migrates=1]	.241	.080	9.004	1	.003	.083	.398
	[Migrates=2]	0(a)	.	.	0	.	.	.
	[Harvests=1]	-.027	.095	.081	1	.776	-.214	.160
	[Harvests=2]	0(a)	.	.	0	.	.	.
	[VGarden=1]	.433	.074	34.621	1	.000	.289	.577
	[VGarden=2]	0(a)	.	.	0	.	.	.
	[LivSource=1]	1.139	.252	20.488	1	.000	.646	1.632
	[LivSource=2]	.807	.250	10.415	1	.001	.317	1.298

[LivSource=3]	.898	.303	8.781	1	.003	.304	1.492
[LivSource=4]	.069	.318	.047	1	.828	-.555	.693
[LivSource=5]	.586	.326	3.238	1	.072	-.052	1.224
[LivSource=6]	.389	.280	1.933	1	.164	-.160	.939
[LivSource=7]	.041	.307	.018	1	.893	-.560	.643
[LivSource=8]	.066	.399	.028	1	.868	-.716	.849
[LivSource=9]	.592	.467	1.604	1	.205	-.324	1.507
[LivSource=10]	.887	.389	5.199	1	.023	.125	1.650
[LivSource=11]	.294	.328	.803	1	.370	-.349	.936
[LivSource=12]	0(a)	.	.	0	.	.	.
[CerealSrce=1]	.865	.625	1.917	1	.166	-.360	2.090
[CerealSrce=2]	1.005	.629	2.557	1	.110	-.227	2.238
[CerealSrce=3]	1.152	.699	2.716	1	.099	-.218	2.522
[CerealSrce=4]	-.362	.774	.219	1	.640	-1.878	1.154
[CerealSrce=5]	.377	1.011	.139	1	.709	-1.605	2.359
[CerealSrce=6]	1.843	.779	5.596	1	.018	.316	3.370
[CerealSrce=7]	.797	.638	1.558	1	.212	-.454	2.048
[CerealSrce=8]	0(a)	.	.	0	.	.	.
[FoodShk=1]	.090	.070	1.648	1	.199	-.047	.226
[FoodShk=2]	0(a)	.	.	0	.	.	.
[FoodAid=1]	-.017	.082	.044	1	.834	-.178	.143
[FoodAid=2]	0(a)	.	.	0	.	.	.
[WIQntile=1]	.105	.121	.751	1	.386	-.132	.341
[WIQntile=2]	.191	.111	2.943	1	.086	-.027	.410
[WIQntile=3]	.126	.106	1.409	1	.235	-.082	.333
[WIQntile=4]	.050	.100	.246	1	.620	-.147	.246
[WIQntile=5]	0(a)	.	.	0	.	.	.

Link function: Cauchit.

a This parameter is set to zero because it is redundant.

Test of Parallel Lines(c)

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Null Hypothesis	5432.723			
General	5307.092(a)	125.630(b)	49	.000

The null hypothesis states that the location parameters (slope coefficients) are the same across response categories.

a The log-likelihood value cannot be further increased after maximum number of step-halving.

b The Chi-Square statistic is computed based on the log-likelihood value of the last iteration of the general model. Validity of the test is uncertain.

c Link function: Cauchit.

APPENDIX 4c

SPSS LINEAR REGRESSION OUTPUTS AND CODE

Regression

```
[DataSet1] E:\UKZN_MSc\DataSetFinal.sav
```

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta	Zero-order	Partial	Part	Tolerance	VIF	B	Std. Error
1	(Constant)	52.982	5.752		9.210	.000					
	State	-.261	.057	-.091	-4.565	.000	-.038	-.086	-.078	.741	1.350
	Type of Household	.677	.612	.019	1.106	.269	-.028	.021	.019	.945	1.058
	Number of household members	.359	.133	.047	2.705	.007	.082	.051	.046	.951	1.052
	Level of education attained	-.423	.391	-.019	-1.082	.279	-.028	-.020	-.019	.946	1.057
	Sex of household head	-.056	1.103	-.001	-.050	.960	.009	-.001	-.001	.957	1.045
	Owned land for agriculture	6.250	1.044	.114	5.987	.000	.146	.112	.102	.801	1.248
	Used land for agriculture	10.460	1.482	.134	7.057	.000	.171	.132	.121	.813	1.230
	Land planted previous year	-.522	.859	-.011	-.607	.544	-.048	-.011	-.010	.867	1.153
	Owned livestock	-6.520	.804	-.147	-8.106	.000	-.159	-.151	-.139	.885	1.130
	Usually migrated	-5.286	.863	-.111	-6.126	.000	-.159	-.115	-.105	.892	1.121
	No. of harvests in one year	3.960	.967	.076	4.095	.000	.098	.077	.070	.850	1.177
	Months food last	1.178	.160	.133	7.374	.000	.167	.138	.126	.898	1.113
	Vegetable garden	-5.780	.794	-.132	-7.283	.000	-.180	-.136	-.125	.889	1.125
	Main sources of livelihood	-.956	.142	-.119	-6.726	.000	-.162	-.126	-.115	.932	1.073
	Meals per day (adults and Children)	3.466	.565	.111	6.136	.000	.127	.115	.105	.898	1.114
	Main source of sorghum and millet	.490	.278	.032	1.759	.079	-.005	.033	.030	.893	1.119
	Experienced food shock	.265	.764	.006	.346	.729	-.016	.007	.006	.958	1.043
	Received food aid	1.045	.841	.023	1.244	.214	.044	.023	.021	.870	1.150
	Wealth index quintiles	-.690	.288	-.042	-2.397	.017	-.065	-.045	-.041	.960	1.042
2	(Constant)	52.914	5.593		9.460	.000					
	State	-.261	.057	-.091	-4.566	.000	-.038	-.086	-.078	.741	1.350
	Type of Household	.675	.611	.019	1.106	.269	-.028	.021	.019	.947	1.056
	Number of household members	.359	.133	.047	2.705	.007	.082	.051	.046	.951	1.052

	Level of education attained	-.419	.384	-.019	-1.092	.275	-.028	-.021	-.019	.980	1.020
	Owned land for agriculture	6.251	1.044	.114	5.990	.000	.146	.112	.102	.802	1.247
	Used land for agriculture	10.457	1.481	.134	7.060	.000	.171	.132	.121	.814	1.229
	Land planted previous year	-.522	.859	-.011	-.607	.544	-.048	-.011	-.010	.867	1.153
	Owned livestock	-6.521	.804	-.147	-8.112	.000	-.159	-.151	-.139	.886	1.129
	Usually migrated	-5.286	.863	-.111	-6.127	.000	-.159	-.115	-.105	.892	1.121
	No. of harvests in one year	3.960	.967	.076	4.096	.000	.098	.077	.070	.850	1.177
	Months food last	1.177	.160	.133	7.379	.000	.167	.138	.126	.900	1.111
	Vegetable garden	-5.779	.793	-.132	-7.285	.000	-.180	-.136	-.125	.889	1.124
	Main sources of livelihood	-.956	.142	-.119	-6.728	.000	-.162	-.126	-.115	.932	1.073
	Meals per day (adults and Children)	3.467	.564	.111	6.143	.000	.127	.115	.105	.899	1.112
	Main source of sorghum and millet	.490	.278	.032	1.761	.078	-.005	.033	.030	.894	1.119
	Experienced food shock	.264	.764	.006	.346	.730	-.016	.007	.006	.959	1.043
	Received food aid	1.045	.840	.023	1.244	.214	.044	.023	.021	.870	1.150
	Wealth index quintiles	-.691	.288	-.042	-2.402	.016	-.065	-.045	-.041	.962	1.040
3	(Constant)	53.160	5.547		9.584	.000					
	State	-.259	.057	-.090	-4.553	.000	-.038	-.086	-.078	.745	1.342
	Type of Household	.658	.609	.019	1.081	.280	-.028	.020	.018	.953	1.049
	Number of household members	.357	.133	.047	2.694	.007	.082	.051	.046	.952	1.050
	Level of education attained	-.421	.384	-.019	-1.098	.272	-.028	-.021	-.019	.980	1.020
	Owned land for agriculture	6.243	1.043	.114	5.985	.000	.146	.112	.102	.802	1.247
	Used land for agriculture	10.469	1.480	.134	7.071	.000	.171	.132	.121	.814	1.229
	Land planted previous year	-.487	.853	-.010	-.570	.568	-.048	-.011	-.010	.880	1.137
	Owned livestock	-6.505	.802	-.147	-8.107	.000	-.159	-.151	-.139	.889	1.125
	Usually migrated	-5.287	.863	-.111	-6.128	.000	-.159	-.115	-.105	.892	1.121
	No. of harvests in one year	3.952	.966	.076	4.090	.000	.098	.077	.070	.850	1.176
	Months food last	1.175	.159	.133	7.372	.000	.167	.138	.126	.902	1.109
	Vegetable garden	-5.776	.793	-.132	-7.283	.000	-.180	-.136	-.125	.889	1.124
	Main sources of livelihood	-.956	.142	-.119	-6.729	.000	-.162	-.126	-.115	.932	1.073
	Meals per day (adults and Children)	3.478	.563	.111	6.174	.000	.127	.116	.106	.902	1.109
	Main source of sorghum and millet	.487	.278	.032	1.750	.080	-.005	.033	.030	.895	1.117
	Received food aid	1.048	.840	.023	1.247	.212	.044	.024	.021	.870	1.150
	Wealth index quintiles	-.693	.287	-.042	-2.411	.016	-.065	-.045	-.041	.962	1.039
4	(Constant)	52.447	5.403		9.707	.000					

	State	-.256	.057	-.089	-4.520	.000	-.038	-.085	-.077	.752	1.329
	Type of Household	.649	.609	.019	1.067	.286	-.028	.020	.018	.954	1.049
	Number of household members	.359	.133	.047	2.707	.007	.082	.051	.046	.953	1.050
	Level of education attained	-.414	.383	-.019	-1.080	.280	-.028	-.020	-.018	.981	1.019
	Owned land for agriculture	6.226	1.043	.114	5.971	.000	.146	.112	.102	.803	1.246
	Used land for agriculture	10.423	1.478	.133	7.052	.000	.171	.132	.121	.816	1.225
	Owned livestock	-6.521	.802	-.147	-8.133	.000	-.159	-.152	-.139	.890	1.123
	Usually migrated	-5.276	.862	-.111	-6.118	.000	-.159	-.115	-.105	.893	1.120
	No. of harvests in one year	3.930	.965	.075	4.070	.000	.098	.077	.070	.851	1.175
	Months food last	1.183	.159	.134	7.457	.000	.167	.139	.128	.910	1.099
	Vegetable garden	-5.841	.785	-.134	-7.443	.000	-.180	-.139	-.127	.908	1.101
	Main sources of livelihood	-.962	.142	-.120	-6.791	.000	-.162	-.127	-.116	.937	1.067
	Meals per day (adults and Children)	3.510	.560	.112	6.262	.000	.127	.117	.107	.911	1.098
	Main source of sorghum and millet	.461	.274	.030	1.679	.093	-.005	.032	.029	.920	1.087
	Received food aid	.996	.835	.022	1.192	.233	.044	.022	.020	.880	1.136
	Wealth index quintiles	-.684	.287	-.042	-2.385	.017	-.065	-.045	-.041	.965	1.037
5	(Constant)	53.006	5.378		9.857	.000					
	State	-.250	.056	-.087	-4.433	.000	-.038	-.083	-.076	.760	1.315
	Number of household members	.363	.133	.048	2.738	.006	.082	.052	.047	.954	1.049
	Level of education attained	-.423	.383	-.019	-1.103	.270	-.028	-.021	-.019	.982	1.018
	Owned land for agriculture	6.094	1.035	.112	5.886	.000	.146	.110	.101	.814	1.228
	Used land for agriculture	10.420	1.478	.133	7.049	.000	.171	.132	.121	.816	1.225
	Owned livestock	-6.532	.802	-.148	-8.147	.000	-.159	-.152	-.139	.890	1.123
	Usually migrated	-5.310	.862	-.111	-6.162	.000	-.159	-.116	-.105	.894	1.118
	No. of harvests in one year	3.891	.965	.075	4.033	.000	.098	.076	.069	.853	1.173
	Months food last	1.173	.158	.133	7.408	.000	.167	.138	.127	.913	1.096
	Vegetable garden	-5.811	.784	-.133	-7.409	.000	-.180	-.138	-.127	.909	1.100
	Main sources of livelihood	-.955	.141	-.119	-6.749	.000	-.162	-.126	-.115	.939	1.065
	Meals per day (adults and Children)	3.479	.560	.111	6.215	.000	.127	.116	.106	.913	1.095
	Main source of sorghum and millet	.484	.273	.031	1.769	.077	-.005	.033	.030	.926	1.080
	Received food aid	1.034	.834	.023	1.239	.215	.044	.023	.021	.882	1.134
	Wealth index quintiles	-.683	.287	-.041	-2.379	.017	-.065	-.045	-.041	.965	1.037
6	(Constant)	52.339	5.344		9.794	.000					
	State	-.249	.056	-.087	-4.424	.000	-.038	-.083	-.076	.760	1.315
	Number of household members	.359	.133	.047	2.711	.007	.082	.051	.046	.954	1.048
	Owned land for agriculture	6.110	1.035	.112	5.902	.000	.146	.111	.101	.814	1.228
	Used land for agriculture	10.379	1.478	.133	7.024	.000	.171	.131	.120	.817	1.224

	Owned livestock	-6.551	.802	-.148	-8.173	.000	-.159	-.152	-.140	.891	1.123
	Usually migrated	-5.296	.862	-.111	-6.146	.000	-.159	-.115	-.105	.894	1.118
	No. of harvests in one year	3.903	.965	.075	4.045	.000	.098	.076	.069	.853	1.173
	Months food last	1.175	.158	.133	7.422	.000	.167	.139	.127	.913	1.096
	Vegetable garden	-5.803	.784	-.133	-7.400	.000	-.180	-.138	-.127	.909	1.100
	Main sources of livelihood	-.958	.141	-.119	-6.770	.000	-.162	-.127	-.116	.939	1.065
	Meals per day (adults and Children)	3.487	.560	.111	6.230	.000	.127	.117	.107	.914	1.095
	Main source of sorghum and millet	.482	.273	.031	1.762	.078	-.005	.033	.030	.926	1.080
	Received food aid	1.010	.834	.022	1.211	.226	.044	.023	.021	.882	1.133
	Wealth index quintiles	-.719	.285	-.044	-2.523	.012	-.065	-.048	-.043	.978	1.023
7	(Constant)	52.686	5.337		9.872	.000					
	State	-.234	.055	-.081	-4.261	.000	-.038	-.080	-.073	.800	1.250
	Number of household members	.367	.132	.048	2.771	.006	.082	.052	.047	.956	1.046
	Owned land for agriculture	6.088	1.035	.111	5.881	.000	.146	.110	.101	.815	1.228
	Used land for agriculture	10.419	1.478	.133	7.051	.000	.171	.132	.121	.817	1.224
	Owned livestock	-6.538	.802	-.148	-8.157	.000	-.159	-.152	-.140	.891	1.122
	Usually migrated	-5.310	.862	-.111	-6.162	.000	-.159	-.115	-.105	.894	1.118
	No. of harvests in one year	4.000	.962	.077	4.160	.000	.098	.078	.071	.859	1.165
	Months food last	1.172	.158	.133	7.403	.000	.167	.138	.127	.913	1.095
	Vegetable garden	-5.856	.783	-.134	-7.477	.000	-.180	-.140	-.128	.912	1.096
	Main sources of livelihood	-.966	.141	-.121	-6.835	.000	-.162	-.128	-.117	.942	1.062
	Meals per day (adults and Children)	3.512	.559	.112	6.278	.000	.127	.118	.107	.915	1.093
	Main source of sorghum and millet	.441	.271	.029	1.625	.104	-.005	.031	.028	.940	1.064
	Wealth index quintiles	-.721	.285	-.044	-2.530	.011	-.065	-.048	-.043	.978	1.023
8	(Constant)	53.883	5.287		10.191	.000					
	State	-.247	.054	-.086	-4.539	.000	-.038	-.085	-.078	.817	1.225
	Number of household members	.379	.132	.050	2.868	.004	.082	.054	.049	.959	1.042
	Owned land for agriculture	6.092	1.035	.112	5.883	.000	.146	.110	.101	.815	1.228
	Used land for agriculture	10.460	1.478	.134	7.078	.000	.171	.132	.121	.817	1.223
	Owned livestock	-6.483	.801	-.147	-8.093	.000	-.159	-.151	-.138	.893	1.120
	Usually migrated	-5.274	.862	-.111	-6.121	.000	-.159	-.115	-.105	.895	1.117
	No. of harvests in one year	4.121	.959	.079	4.297	.000	.098	.081	.074	.864	1.158
	Months food last	1.158	.158	.131	7.323	.000	.167	.137	.125	.916	1.092
	Vegetable garden	-5.810	.783	-.133	-7.421	.000	-.180	-.139	-.127	.913	1.095
	Main sources of livelihood	-.932	.140	-.116	-6.666	.000	-.162	-.125	-.114	.963	1.039
	Meals per day (adults and Children)	3.513	.560	.112	6.279	.000	.127	.118	.107	.915	1.093
	Wealth index quintiles	-.700	.285	-.042	-2.456	.014	-.065	-.046	-.042	.980	1.020

a Dependent Variable: Food Consumption Score

Excluded Variables(h)

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
		Tolerance	VIF	Minimum Tolerance	Tolerance	VIF	Minimum Tolerance	Tolerance
2	Sex of household head	-.001(a)	-.050	.960	-.001	.957	1.045	.741
3	Sex of household head	-.001(b)	-.045	.964	-.001	.957	1.045	.745
	Experienced food shock	.006(b)	.346	.730	.007	.959	1.043	.741
4	Sex of household head	-.001(c)	-.042	.967	-.001	.957	1.045	.752
	Experienced food shock	.005(c)	.276	.782	.005	.972	1.029	.749
	Land planted previous year	-.010(c)	-.570	.568	-.011	.880	1.137	.745
5	Sex of household head	.000(d)	.005	.996	.000	.959	1.043	.760
	Experienced food shock	.003(d)	.192	.848	.004	.978	1.022	.758
	Land planted previous year	-.010(d)	-.543	.587	-.010	.880	1.136	.753
	Type of Household	.019(d)	1.067	.286	.020	.954	1.049	.752
6	Sex of household head	.004(e)	.212	.832	.004	.994	1.006	.760
	Experienced food shock	.004(e)	.212	.832	.004	.978	1.022	.758
	Land planted previous year	-.009(e)	-.507	.612	-.010	.881	1.135	.753
	Type of Household	.019(e)	1.090	.276	.021	.954	1.048	.752
	Level of education attained	-.019(e)	-1.103	.270	-.021	.982	1.018	.760
7	Sex of household head	.004(f)	.206	.837	.004	.994	1.006	.800
	Experienced food shock	.004(f)	.234	.815	.004	.979	1.022	.797
	Land planted previous year	-.007(f)	-.371	.710	-.007	.892	1.121	.796
	Type of Household	.020(f)	1.141	.254	.022	.956	1.046	.789
	Level of education attained	-.018(f)	-1.072	.284	-.020	.983	1.018	.800
	Received food aid	.022(f)	1.211	.226	.023	.882	1.133	.760
8	Sex of household head	.003(g)	.177	.859	.003	.994	1.006	.814
	Experienced food shock	.003(g)	.193	.847	.004	.979	1.021	.813
	Land planted previous year	-.002(g)	-.117	.907	-.002	.913	1.095	.809
	Type of Household	.022(g)	1.258	.209	.024	.961	1.040	.803
	Level of education attained	-.018(g)	-1.066	.287	-.020	.983	1.018	.814
	Received food aid	.018(g)	1.003	.316	.019	.896	1.116	.770
	Main source of sorghum and millet	.029(g)	1.625	.104	.031	.940	1.064	.800

APPENDIX 5

SIGNIFICANT PREDICTORS FOR FITTING A FINAL PREDICTED MODEL

	Predictor and Levels		Odds Ratio Estimate	Confidence Interval		Significance Probability
				Lower	Upper	
1	State	Jonglei	2.74	1.91	3.92	0.000000
		Upper Nile	3.12	2.19	4.44	0.000000
		Unity	4.40	3.03	6.40	0.000000
		Warrap	1.77	1.27	2.47	0.000796
		Northern Bahr El Ghazal	1.52	1.10	2.12	0.012400
		Western Bahr El Ghazal	2.39	1.73	3.30	0.000000
		Lakes	1.85	1.39	2.46	0.000025
		Western Equatoria	1.68	1.26	2.24	0.000376
		Central Equatoria	3.98	2.90	5.45	0.000000
2	Months of harvest food lasting		1.09	1.06	1.13	0.00000
3	Number of meals per day		1.41	1.26	1.58	0.00000
4	Farmland ownership		0.83	0.68	1.01	0.060364
5	Farmland use		0.58	0.43	0.77	0.000225
6	Farmland planting		1.24	1.06	1.46	0.006581
7	Livestock ownership		1.47	1.26	1.71	0.000001
8	Migration/movement of household		1.27	1.09	1.49	0.002693
9	Ownership of home garden		1.54	1.33	1.78	0.000000

	Predictor and Levels		Odds Ratio Estimate	Confidence Interval		Significance Probability*
				Lower	Upper	
10	Source of livelihoods	Livestock rearing	3.12	1.91	5.11	0.000006
		Agricultural production	2.24	1.37	3.66	0.001250
		Fishing	2.45	1.36	4.44	0.003044
		Hunting, gathering	1.07	0.57	2.00	**0.828124**
		Petty trade	1.80	0.95	3.40	0.071950
		Collecting natural resources	1.48	0.85	2.56	**0.164426**
		Unskilled labour	1.04	0.57	1.90	**0.892877**
		Handicrafts	1.07	0.49	2.34	**0.867961**
		Skilled labour	1.81	0.72	4.51	**0.205398**
		Employed work	2.43	1.13	5.20	0.022597
		Food aid assistance	1.34	0.71	2.55	**0.370176**
11	Source of sorghum and millet	Own production	2.38	0.70	8.09	**0.166231**
		Market purchase	2.73	0.80	9.37	**0.109840**
		Hunting/fishing/gathering	3.16	0.80	12.46	**0.099373**
		Exchange	0.70	0.15	3.17	**0.639840**
		Borrowed	1.46	0.20	10.58	**0.709136**
		Gift	6.32	1.37	29.09	0.018004
		Food aid	2.22	0.63	7.75	**0.212026**

* Non-significant probabilities are bolded.