

Addis Ababa  
University  
(Since 1950)



**Addis Ababa University  
Postgraduate Program**

# **DETERMINANTS OF FOOD SECURITY IN RURAL HOUSEHOLDS OF THE TIGRAY REGION**

**BY:  
TSEGAY GEBREHIWOT**

**June 2009**

**Addis Ababa University  
Postgraduate Program**

**DETERMINANTS OF FOOD SECURITY IN RURAL HOUSEHOLDS OF  
THE TIGRAY REGION**

**BY:  
TSEGAY GEBREHIWOT**

**A THESIS SUBMITTED TO SCHOOL OF GRADUATE STUDIES OF ADDIS ABABA  
UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE  
DEGREE OF MASTER OF SCIENCE IN APPLIED STATISTICS**

**June 2009**

**Addis Ababa University  
Postgraduate program**

**DETERMINANTS OF FOOD SECURITY IN RURAL HOUSEHOLDS OF  
THE TIFRAY REGION**

**BY:  
TSEGAY GEBREHIWOT**

**FACULTY OF SCIENCE  
DEPARTMENT OF STATISTICS**

**Approved by the Examining Board**

|  |                    |
|--|--------------------|
| .....<br>Chairman, Department Graduate Committee | .....<br>Signature |
| .....<br>Advisor                                 | .....<br>Signature |
| .....<br>Internal Examiner                       | .....<br>Signature |
| .....<br>External Examiner                       | .....<br>Signature |

## Acknowledgement

I owe the deepest gratitude to Dr. Emmanuel G/Yohannes, my thesis adviser and my instructor, for his invaluable and constructive comments and suggestions throughout my study. Without his professional assistance and guidance this study would not be realized.

I would like to acknowledge the employees of CSA, especially W/o Elleni for making the data available and Ato Zenaselassie for giving me enough information about the data.

I would also like to thank my employer, University of Gondar, for the sponsorship that allowed me to study my postgraduate education.

Finally, this might be a good opportunity to acknowledge all members of my family, relatives and friends for all their encouragement and support throughout my study.

# Table of Contents

| Contents  | Page |
|---|------|
| Acknowledgement   | i    |
| Table of Contents   | ii   |
| List of Tables  | iv   |
| List of Figures   | iv   |
| List of Abbreviations   | v    |
| ABSTRACT  | vi   |
| CHAPTER ONE: INTRODUCTION   | 1    |
| 1.1 General Background  | 1    |
| 1.2 General Situation of the Study area                                 | 2    |
| 1.3 Statement of the problem  | 4    |
| 1.4 Objective of the study  | 5    |
| 1.5 Limitation of the Study   | 6    |
| CHAPTER TWO: LITERATURE REVIEW  | 7    |
| 2.1 Conceptual Framework  | 7    |
| 2.1.1 Concepts and Definitions of Food Security                         | 7    |
| 2.1.2 Theoretical Approaches to Food Security                           | 9    |
| 2.1.2.1 The Food Availability Decline (FAD) Approach                    | 10   |
| 2.1.2.2 The Food Entitlement Decline (FED) Approach                     | 10   |
| 2.2. Empirical Review of Causes and Determinants of Food Insecurity     | 11   |
| 2.2.1 Causes of Food Insecurity   | 11   |
| 2.2.1.1 Causes of Food Insecurity in Other Developing Countries         | 11   |
| 2.2.1.2. The Ethiopian Case   | 13   |
| 2.2.2. Determinants of household food security                          | 14   |
| 2.2.3 Generalizations of the Causes and Determinants of Food Insecurity | 15   |
| 2.2.3.1. Demographic Factors  | 15   |
| 2.2.3.2. Environmental Crises   | 16   |
| 2.2.3.3. Poor Asset Base of the Rural Households                        | 16   |
| 2.2.3.4. Socio-Cultural Factors   | 17   |
| 2.2.3.5. War or Conflict  | 17   |
| 2.2.3.6. Access to Infrastructure                                       | 18   |
| 2.3. Comments on the Reviewed Literatures                               | 18   |
| CAPTER THREE: METHODOLOGY   | 20   |
| 3.1 Data Source   | 20   |
| 3.2 Variables in the study  | 22   |
| 3.2.1 The Dependent variable  | 22   |
| 3.2.1.1 Measurement of food insecurity                                  | 23   |
| 3.2.2 Explanatory Variables   | 24   |
| 3.2.2.1. Description of Explanatory Variables                           | 25   |
| 3.3 Model Selection   | 28   |
| 3.4 Discriminant Analysis   | 29   |
| 3.4.1. Assumptions of Discriminant Analysis                             | 30   |

|   |           |
|---|-----------|
| 3.4.2 Theoretical Development of Discrimination and Classification..... | 30        |
| 3.4.2.1 Separation and classification for two populations .....         | 31        |
| 3.4.2.2 Optimal Classification rules for two populations .....          | 33        |
| 3.4.2.3 Classification with two multivariate normal populations .....   | 34        |
| 3.4.3 Applications of Discriminant and Classification Analysis .....    | 36        |
| 3.4.3.1 The Discriminant Function for Two Groups .....                  | 38        |
| 3.4.3.2 Standardized Discriminant Function.....                         | 40        |
| 3.4.3.3 Tests of Significance in Discriminant analysis.....             | 40        |
| 3.4.3.3.1 Test of equality of means between two groups.....             | 41        |
| 3.4.3.3.2 Test of Equality of Covariance Matrices of Two Groups.....    | 42        |
| 3.4.3.3.3 Measure of Association for Discriminant Function.....         | 43        |
| 3.4.3.3.4 Test of Significance of the Discriminant function.....        | 43        |
| 3.4.3.4 Stepwise Selection of Variables.....                            | 43        |
| 3.4.3.5 Classification.....   | 45        |
| 3.4.3.6 Estimating Misclassification rates .....                        | 46        |
| <b>CHAPTER FOUR.....</b>  | <b>47</b> |
| <b>SATATISTICAL DATA ANALYSIS .....</b>                                 | <b>47</b> |
| 4.1 Evaluation of Assumptions .....                                     | 47        |
| 4.1.1 Checking Multivariate Normality .....                             | 47        |
| 4.1.2 Testing Homogeneity of Covariance Matrices between Groups .....   | 48        |
| 4.1.3 Checking outliers .....   | 49        |
| 4.1.4 Checking Multicollinearity .....                                  | 49        |
| 4.2 Descriptive Analysis Results .....                                  | 50        |
| 4.3 Bivariate Analysis Results .....                                    | 52        |
| 4.4 Results of stepwise variable selection.....                         | 55        |
| 4.5 Results of Discriminant Analysis and Discussion .....               | 56        |
| 4.5.1 Test of Equality of group means.....                              | 56        |
| 4.5.2 Significance of the Discriminant Function .....                   | 58        |
| 4.5.3 Discriminant Function and its Coefficients .....                  | 59        |
| 4.5.2.1 Unstandardized Discriminant Function Coefficients.....          | 60        |
| 4.5.2.2 Standardized Discriminant Function Coefficients.....            | 61        |
| 4.5.4 Structure Matrix.....   | 65        |
| 4.5.5 Group Centroids.....  | 67        |
| 4.6 Classification results .....  | 68        |
| <b>CHAPTER FIVE .....</b>   | <b>72</b> |
| <b>CONCLUSIONS AND RECOMMENDATIONS .....</b>                            | <b>72</b> |
| 5.1 Conclusions.....  | 72        |
| 5.2 Recommendations.....  | 74        |
| References.....   | 77        |
| Annexes.....  | 83        |
| ANNEX A: Tables of conversion factors .....                             | 83        |
| ANNEX B: Test of Assumptions.....                                       | 85        |
| ANNEX C: Normal Q-Q plots, Chi-square plots and Box plot.....           | 86        |
| ANNEX D: Casewise Statistics .....                                      | 89        |

| <b>List of Tables</b>   | page |
|---|------|
| Table 3.1: Demographic variables.....   | 24   |
| Table 3.2: Socio-Economic Variables.....  | 24   |
| Table 4.1: Box’s Test of Equality of Covariance Matrices.....                   | 48   |
| Table 4.2: Food Insecurity Indices.....   | 51   |
| Table 4.3: Group Statistics for continuous variables.....                       | 51   |
| Table 4.4: Association between HFS status and predictor variables .....         | 54   |
| Table 4.5: Variables selected by stepwise procedure.....                        | 55   |
| Table 4.6: MANOVA for the Test of Equality of group means.....                  | 56   |
| Table 4.7: Univariate ANOVA for the Test of Equality of Group Means.....        | 57   |
| Table 4.8: Eigenvalues.....   | 58   |
| Table 4.9: Wilks’ lambda.....   | 59   |
| Table 4.10: Discriminant Function Coefficients.....                             | 62   |
| Table 4.11: Structure matrix.....   | 66   |
| Table 4.12: Functions at group Centroids.....                                   | 67   |
| Table 4.13: Prior Probabilities for Groups.....                                 | 68   |
| Table 4.14: Classification Function Coefficients.....                           | 69   |
| Table 4.15: Classification Results.....   | 70   |
| Table A1: Conversion factor used to estimate Tropical livestock Unit (TLU)..... | 83   |
| Table A2: Conversion factor used to calculate adult equivalence scales.....     | 83   |
| Table A3: Caloric content of the food consumed in the study area.....           | 84   |
| Table B1: Collinearity information of predictor variables.....                  | 85   |
| Table B2: Coefficients of Skewness and Kurtosis for continuous variables.....   | 85   |
| Table D1: Case statistics.....  | 89   |

| <b>List of Figures</b>   | page               |
|--|--------------------|
| Figure 2.1: Food Security status Framework.....  | 19                 |
| Figure C1: Normal Q-Q plot of discriminant discriminant scores for all households..... | 86                 |
| Figure C2: Chi-square plot for all household.....                                      | 86                 |
| Figure C3: Chi-square plot for food secure households.....                             | 87                 |
| Figure C4: Chi-square plot for food insecure households.....                           | <a href="#">87</a> |
| Figure 4.1: Boxplot of discriminant scores.....  | 88                 |

## List of Abbreviations

|        |  |
|--------|--|
| ANOVA  | Analysis of Variance                                 |
| CDR    | Child Dependency Ratio                               |
| CSA    | Central Statistical Agency                           |
| DA     | Discriminant Analysis                                |
| ECM    | Expected Cost of Misclassification                   |
| EEA    | Ethiopian Economic Association                       |
| FAO    | Food and Agricultural Organization                   |
| GDP    | Gross Domestic Product                               |
| GAO    | Government Accountability Office (USA)               |
| HFS    | Household Food Security                              |
| HH     | Household  |
| HICES  | Household Income, Consumption and Expenditure Survey |
| IFPRI  | International Food Policy Research Institute         |
| Kcal   | Kilo calorie   |
| LDA    | Linear Discriminant Analysis                         |
| MEDaC  | Ministry of Economic Development and Cooperation     |
| MANOVA | Multivariate Analysis of Variance                    |
| MoFED  | Ministry of Finance and Economic Development         |
| REST   | Relief Society of Tigray                             |
| TLU    | Tropical Livestock Unit                              |
| SPSS   | Statistical Package for Social Science               |
| Q-Q    | Quantile-quantile                                    |
| UNDP   | United Nations development Program                   |
| VIF    | Variance Inflation Factor                            |
| WFS    | World Food Summit                                    |
| WMS    | Welfare Monitoring Survey                            |

## ABSTRACT

This study investigated the determinants of food security and identified the major factors that jointly discriminate the rural households of Tigray region into food secure and food insecure households. The study is made based on the 2004/2005 Household Income, Consumption and Expenditure Survey (HICES) and Welfare Monitoring Survey (WMS) which were conducted by Central Statistical Agency (CSA). To analyze the data descriptive statistics, bivariate and multivariate analyses were used. The descriptive results revealed that about 42% of the households were found to be food insecure while 57% were food secure. The bivariate analysis was performed to investigate the effect of each predictor variable on the household food security status. A priori expectations about the relationships between household food security status and factors influencing it were satisfied for the majority of the cases considered. Moreover, a univariate ANOVA of each predictor variables against the household food security was performed to identify the variables that have significant contribution to the discrimination of the two household groups. Accordingly, distance to input sources, farmland size, TLU, number of oxen, household size were found to be the major discriminating variables. This was further supported by multivariate discriminant function analysis applied to sampled farm households. The importance of the contribution of factors in discriminating the two household groups were ranked by the discriminant function. As a result, distance to input source was ranked first followed by household size, farmland size, livestock ownership, number of oxen, use of fertilizer, gender and educational level of the household head. In addition to this the classification results revealed that 79.7% of the households were correctly classified. This indicates that the discriminant function employed in this study was efficient in discriminating the rural households based on the eight predictors.

# CHAPTER ONE: INTRODUCTION

## 1.1 General Background

We are living in a world where more than one billion people are poor, 800 million are food insecure, and where about 170 million children are malnourished. While food insecurity occurs in most countries to varying degrees, 75 % of the food insecure lives in rural areas of developing countries (FAO, 2002; IFPRI, 2002). Food is both a need and human right, but food insecurity is prevalent in today's world in general, and in sub-Saharan Africa in particular. Since early 2007, food-related riots have occurred in 15 countries, including 7 in sub-Saharan Africa (GAO, 2008).

Africa, which reversed from being a key exporter of agricultural commodities into being a net importer, has the highest percentage of undernourished and has shown the least progress on reducing the prevalence of undernourishment in the last 30 years. Chronic food insecurity now affects about 200 million people who are suffering from malnutrition. Acute food insecurity in 2003 affected 38 million people in Africa who are facing the outright risk of famine, with 24,000 dying from hunger daily. Famines are the most visible and extreme manifestation of acute food insecurity. Out of the 39 countries worldwide that faced food emergencies at the beginning of 2003, 25 are found in Africa including Ethiopia (Clover, 2003).

As part of Africa, Ethiopia faces daunting poverty and food insecurity challenges that are worsening over time. About half of Africa's food insecure population lives in Ethiopia, Chad, Zaire, Uganda, Zambia and Somalia (Ramakrishna *et al.*, 2002). In the 1990s, an estimated 30 million Ethiopians were food insecure, and food crises were persistent. Among this food insecure people the majority reside in the rural areas of the country. About 52% of the rural population and 36% of the urban population consume under the minimum recommended daily intake of 2100 calorie per person per day (FAO, 2002; MEDaC, 1999). The world development report indicators for the year 2000/01 reveal the prevalence of child malnutrition (children under age 5) is 48% during the period 1992-1998 ( Berhanu, 2004).

Ethiopia has reasonably good resource potential for development –agriculture, biodiversity, water resource, minerals, etc. Yet, Ethiopia is faced with complex poverty, which is broad, deep and structural. The proportion of the population below the poverty line is 44% in 1999/2000 (MoFED, 2002).

The presence of hunger in Ethiopian households due to insufficient resources to obtain food has been a long-standing challenge to Ethiopian government, donors, and other local and international organizations. Despite significant amounts of food aid assistance over recent years, there has been little measurable impact in reducing food insecurity. The reason behind is that food insecurity is a complex, multidimensional phenomenon which varies through continuum of successive stages as the condition becomes more severe. Each stage consists of characteristic conditions and experiences of food insufficiency to fully meet the basic needs of household members, and of the behavioral responses of household members to these conditions (USAID, 2008).

## **1.2 General Situation of the Study area**

The Tigray National Regional State (the study area) is located in the most Northern part of Ethiopia. It is bordered by Afar region in the East, Eritrea in the North, Amhara region in the South and Sudan in the West. Based on the 2007 census conducted by the Central Statistical Agency (CSA), Tigray has an approximate area 53,386 Square-kilometers (about 7% of Ethiopia) and an estimated population of 4,314,456 (about 6% of the country's population) of which 49.2% of the population are males and 50.8% are females. The population grows by 2.5% per year.

This region has an estimated density of 86.15 people per square kilometer. In the entire region, there are about 985,654 households, which results in an average for the region of 4.4 persons per household, with urban households having on average 3.4 and rural households 4.6 persons (CSA, 2008).

Ethnically the Region is predominantly Tigriyan (96.55% of the population); other ethnic groups include Amhara (1.63%), Irob or Saho (0.71%), Afar (0.29%), Agaw (0.19%), Oromo (0.17%), and Kunama (0.07%). 95.6% of the population are Orthodox Christian, 4.0% Muslim, 0.4% Catholic, and 0.1% Protestant (CSA, 2008).

According to the new administrative set up, Tigray is divided into six zones. These are Western, North-Western, Central, Eastern, Southern and Mekelle (the regional capital).

Agro-climatically, it is classified into *Kola* (low land), *Weina dega* (mid high land) and *Dega* (high land) in which the average temperature ranges from 4<sup>0</sup>c to 40<sup>0</sup>c with some exceptions. The rainfall of the region varies from 450mm to 900mm which makes the region usually moisture deficient resulting in recurrent drought. The soils of Tigray consist of predominantly litho sols, red clay loam, nitro sols, and sandy clay loams with characteristic low organic material content, high erodibility and low moisture holding capacity (Yared, 2000).

The topography of the region is characterized by mountain plateau. The mountains vary in altitude from 2000-3000 meters above sea level. The western plateau comprises mostly lowland areas with depressions in the boundaries of the Afar region. One of the notable physical features of Tigray is hills and valleys.

In Tigray above 80% of the population live in the rural areas and is engaged in agriculture. *Kiremt* (summer) is the main rainy season. The peak agricultural season is from June to August while the slack period is from December to April.

The food supply of the region is primarily dependent on Meher production. The major types of crops growing in Tigray are barely, wheat, teff, sorghum, maize, millet, and pulses. The short maturing meher crops such as barley, wheat, and teff are the main crops grown in the high and mid-altitude zones while sorghum and millet are the dominant crops in the low land areas especially in the central and western zones of Tigray. *Belg* rains are vital for *meher* season land preparation and planting of long cycle crops.

Livelihoods in Tigray region are significantly dependent on agricultural income options supported by off-farm income generation such as labor trading and petty trade. However, agricultural production and diversification remain low. In this region the average size of land available to a four persons is about 0.5 hectare, too small to support the family on agricultural production alone. The average production of cereals (the major agricultural output) is less than 7 quintals per

household in the drought prone areas and this level of staple cereal production can only feed a family only for 5-8 months a year at best (CSA, 2005; Kidane, 2006).

Tigray is one of the most disaster prone and food insecure regions in Ethiopia. The region has been dependent on relief assistance for many years. According to the socio-economic survey conducted in the region in 1995, 16 % of the population revealed to be self-supporting, while the vast majority of 84% couldn't support themselves (REST, 1995). This and other similar studies show that there is a problem of chronic vulnerability even in the absence of major drought and insect infestation and in seasons with favorable climatic condition.

The main causes of structural food deficiency in the region include severe environmental degradation, low soil fertility, inadequate and erratic rain fall, vulnerability to pests, lack of appropriate technology, small size and fragmentation of land holdings, lack of diversification, lack of oxen for draft power and little use of modern inputs ( Fasil *et al.*, 2007).

### **1.3 Statement of the problem**

Ethiopia had been faced with many droughts and many people had died of famine than other problems particularly in the epidemic periods of 1957-58, 1964-65, 1983-84, 1998-99, and 2003 (EEA, 2005). Since the last major famine of 1984/5 (when excess mortality may have reached one million), Ethiopia has been affected by recurrent droughts (Devereux, 2000). Some droughts were exacerbated by civil conflict, which undermined food production and inhibited government and donor responses to the harvest failure. In Ethiopia, the seriousness of the famine and food shortage varies from one area to another on the state of natural resources and the extent of development of these resources.

Most famines and food crises have been geographically concentrated along two broad belts of the country. The first belt consists of the mixed farming production system area of the central and northern highlands, stretching from northern Shewa through Wollo and Tigray. The land resources, mainly the soils and vegetation of this part of the country have been highly degraded because of the interplay between some environmental and human factors such as relief, climate, population pressure and the resultant over-cultivation of the land, deforestation of vegetation and overgrazing.

The second belt is made up of the low-lying agro-pastoral lands ranging from Wollo in the North, through Hararghe and Bale to Sidamo and Gamo Gofa in the South (Degafa, 2002; Ramakrishna and Assefa, 2002).

The study area, Tigray region, is one of the food deficient regions of the country, which falls in the first drought prone belt. As a result of the food deficient situation in the region, where even in a good year farm households can only meet 60% of their total food needs and the remaining is filled by food aid -both free and Food-For-Work (FARM Africa, 1998; Sosina and H.Stein, 2007).

Although the seriousness of food shortage varied from year to year, farm households faced seasonal food shortage almost every year. Food secure and food insecure farm households reside as neighbors and could share common climate and weather situation and mainly similar socio-economic, cultural and land topography. Yet, one faces seasonal food crises and become dependent on food aid, while the other remains food secure, requiring no food aid. This clearly shows poverty and transitory food insecurity are deep-rooted in the study area. Although drought plays a paramount role in triggering food crises, the difference in consumption status of farm households between good year and bad year is not so significant to claim that drought is the central cause of famine or transitory food insecurity. This implies poverty and seasonal food insecurity in the region are mainly determined by structural, socio-economic, cultural, demographic and other factors. Hence, the central question of this study is what factorial differences make the farm household food secure or food insecure.

## 1.4 Objective of the study

The **main objective** of this study is to assess the status of food security and its major determinants in the rural households of the Tigray National Regional State.

The **specific objectives** of this study include:

1. To identify and evaluate the major factors that affect the status of food security of rural households in Tigray region;

2. to develop a model that predicts the status of rural household's food security on the knowledge of some determining factors;
3. to apply discriminant analysis in classifying rural households of Tigray region based on their status of food security.

## **1.5 Limitation of the Study**

Some of the limitations of this study are:

1. The study focused on identifying factors that are expected to influence household food security in the rural areas of the Tigray regional state. However, due to lack of data the study could not incorporate some of the most influencing factors such as political, climatic and weather (rainfall, temperature); topography, natural disasters and ecological conditions;
2. The study did not made a comparative analysis of food security between rural and urban households in the region;
3. The study was concerned about transitory food insecurity faced by farm households for any magnitude and hence did not deal with causes of chronic food insecurity;
4. Due to lack of recent data on households' food security of the region, the study used data taken from both HICE and WM surveys which were conducted in 2004/5 by CSA. Hence, the result of this study may not correctly reflect the current situation of food security in the study area.

# CHAPTER TWO: LITERATURE REVIEW

## 2.1 Conceptual Framework

### 2.1.1 Concepts and Definitions of Food Security

Since the World Food Conference in 1974 due to food crises and major famines in the world, the term Food Security was introduced, evolved, developed and diversified by different researchers. Maxwell and Frankenberger (1992) listed 194 different studies on the concept and definition of Food Security and 172 studies on indicators. A review that updates this literature (Clay, 1997) provides an additional 72 references. In the work by Maxwell and Frankenberger, a distinction is made between process indicators (those that describe food supply and food access) and outcome indicators (those describe food consumption).

Food security was understood as adequacy of food supply at global and national levels until the mid 1970's. This view favored merely food production oriented variables and overlooked the multiple forces which in many ways affect food access. Evidences show that during the last two decades, food production has been increasing in the world. However, large amount of food at global level does not guarantee food security at national level. Moreover, availability of enough food at national level does not necessarily ensure household food security. For instance, in 1990, the calorie supply at global level was more than 110 percent compared to the total requirement. However, during the same period, more than 100 million people were affected by famine and more than a quarter of the world's population was short of enough food (UNDP, 1992). Although food production has been increasing from time to time, food insecurity, malnutrition and hunger and much more serious problems would remain the main agenda in the globe today (Barrett, 2002).

As the occurrence of hunger, famine, and malnutrition are increasing from time to time in developing countries, the conceptual framework of food security has also progressively developed and expanded. The idea of food security attained wider attention since the 1980s after the debate on 'access' to food and the focus of the unit shifted from global and national levels to household and individual levels (Debebe *et al.*, 1995). This paradigm came with new concept and definition of food security and it led to two additional major shifts in thinking; from a first food approach to a

livelihood perspective and from objective indicators to subjective perceptions (Maxwell *et al.*, 1994).

The most commonly accepted definition of Food security is “*access by all people at all times to enough food for an active and healthy life*” (World Bank, 1986). Food insecurity is a *situation in which individuals have neither physical nor economical access to the nourishment they need*. A household is said to be food insecure when its consumption falls to less than 80% of the daily minimum recommended allowance of caloric intake for an individual to be active and healthy. In particular, food insecurity includes low food intake, variable access to food, and vulnerability- a livelihood strategy that generates adequate food in good times but is not resilient against shocks. These outcomes correspond broadly to chronic, cyclical, and transitory food insecurity, and all are endemic in Ethiopia (Devereux, 2000).

During the debates that preceded the World Food Summit (WFS) held in Rome in 1996, it was established that "*There is food security when all people at all times have sufficient physical and economic access to safe and nutritious food to meet their dietary needs including food preferences, in order to live a healthy and active life*"(USAID,2008). When an individual or population lacks, or is potentially vulnerable due to the absence of, one or more factors outlined above, then it suffers from, or is at risk of, food insecurity. Based on the WFS (1996), the definition focuses on three distinct but interrelated elements, all three of which are essential to achieving food security:

- **Food availability:** having sufficient quantities of food from household production, other domestic output, commercial imports or food assistance,
- **Food access:** having adequate resource to obtain appropriate foods for a nutritious diet, which depends on available income, distribution of income in the household and food prices,
- **Food utilization:** proper biological use of food, requiring a diet with sufficient energy and essential nutrients, potable water and adequate sanitation, as well as knowledge of food storage, processing, basic nutrition and child care and illness management.

The concept of food security also has spatial and temporal dimensions. The spatial dimension refers to the degree of aggregation at which food security is being considered. It is possible to

analyze food security at the global, continental, national, sub-national, village, household, or individual level (Hoddinott, 1999).

The temporal dimension refers to the time frame over which food security is being considered. In much of the food security literature, temporal dimension is almost universally classified in to two states-chronic or transitory (Hoddinott, 1999; Tweeten, 1997; Devereux, 2006). ***Chronic food insecurity*** is a long-term or persistent inability to meet minimum food consumption requirements; while ***transitory food insecurity*** is a short-term or temporary food deficiency. An intermediate category is ***cyclical food insecurity***, such as seasonality. Transitory is often used to imply acute, with the corollary assumption that chronic equates to mild or moderate food insecurity (Devereux, 2006).

The worst form of transitory food insecurity is famine (Devereux, 2006). Hence, transitory food insecurity faced by farm households will be understood in this study as a seasonal food shortage of any magnitude ranging from mild to severe. It can also be noted that the concepts of transitory food insecurity and seasonal food shortages are synonymous and will be used interchangeably in this study. As the Ethiopian farming system is mainly dependent on rain-fed agriculture, seasonality adversely affects the food security situation of the country.

### **2.1.2 Theoretical Approaches to Food Security**

The **general approach** has pointed out a number of environmental and socio-economic attributes assumed to explain famine and food security. The principal ones include: rapid population growth, war and civil strife, drought, ecological degradation, government mismanagement, unequal access to resources and unequal exchange, socio-economic and political dislocation (Getachew, 1995). The argument of this approach is that one or a combination of these can disrupt food production. However, production failure may or may not result in famine or food insecurity. Due to this fact, the attributes (factors) are not precise explanations of the causation of the process of famine. It is in response to this major problem weakness that the specific approaches (models) of famine emerged (Degafa, 2002).

### **2.1.2.1 The Food Availability Decline (FAD) Approach**

The Food Availability Decline Approach had been a dominant theoretical explanatory framework for food crises since the eighteenth century until the year 1980. As quoted in Getachew (1995), Sen (1980) defined FAD as “The availability decline per capita of food for consuming unit”. This approach conceived famine as shortages of food supplies per capita, motivated by natural factors; e.g., drought, floods and other calamities that undermine crops; or demographic factors, i.e., vegetative growth that goes beyond supply (Hewitt, 1993: cited in Diana, 2007).

The central argument of this model is that “anything which disrupts food production such as drought, flood or war can cause famine, the logic being that a drought, flood or war causes crop failure and cattle death, reducing the availability of food in the affected region, and that such a food availability decline for an extended period by definition constitutes of famine” (Markos, 1997; Devereux, 1988: as cited in Degafa, 2002). Hence, this argument claimed that hunger and famine do not necessarily evolve from lack of food supplies in the market, but lack of resources in sectors to produce or purchase them. This criticism over FAD ended up in the alternative model of ‘Entitlement’ proposed by the economist Amartya Sen in 1981.

### **2.1.2.2 The Food Entitlement Decline (FED) Approach**

Amartya Sen’s influential book ‘poverty and Famine’ (1981) decisively shifted the focus of famine analysis from supply side to the demand side. The entitlement approach emphasizes access to food, or people’s relationship to the food, rather than the availability of food (Devereux and Maxwell, 2003). The main argument of this model is the mere presence of food in the economy or in the market does not entitle a person to consume it and thus starvation can set in without any obvious aggregate available fall (Getachew, 1995).

Some of the catastrophic famines have occurred without FAD. For example, the Bengal famine of 1943, the Ethiopian famine of 1973 and 1984, and the Bangladesh famine of 1974 occurred due to lack of entitlement rather than due to lack of availability short fall (Fasil, 2005).

Among many positive features of the FED approach over FAD, the following are very important:

First it has emphasized upon demand rather than supply. Second, it allows vulnerable groups to be identified. Finally, it suggests more appropriate policy intervention (Devereux and Maxwell, 2003).

Although this approach has the above mentioned strength upon FAD, it has also its own limitations. Generally, food security signifies the combination of the above two approaches and food utilization because enough food must be available, and households must have the capabilities to acquire it (Degafa, 2002).

## **2.2. Empirical Review of Causes and Determinants of Food Insecurity**

The empirical review for this study is organized under three sections. The first section presents some causes of food insecurity documented in Ethiopia and other developing countries of the world particularly in Africa. The second part presents determinants of food security in Ethiopia. The last part presents and generalizes the findings of certain previous studies concerning the determinants of food insecurity.

### **2.2.1 Causes of Food Insecurity**

#### **2.2.1.1 Causes of Food Insecurity in Other Developing Countries**

Achieving food security in its totality continues to be a challenge not only for the developing nations, but also for the developed world. The difference lies in the magnitude of the problem in terms of its severity and proportion of the population affected.

Mwanki (2005) mentioned the main causes of food insecurity in developing countries. Some of them include: unstable social and political environments that preclude sustainable economic growth, war and civil strife, macro-economic imbalances in trade, natural resource constraints, poor human resource base, gender inequality, inadequate education, poor health, natural disasters, such as floods and locust infestation, and the absence of good governance. All these factors contribute to either insufficient national food availability or insufficient access to food by households and individuals.

A study by Boussard *et al.* (2005) found that 99% of the food in Sub-Saharan Africa is grown under rain fed agriculture. Hence, food production is vulnerable to adverse weather conditions. The reason behind is that there was an over decline in farm input investment including fertilizers, seeds, and technology adoptions.

Other causes include rapid population growth, limited access to agriculture-related technical assistance, underdeveloped agricultural sector and lack of knowledge about profitable soil fertility management practices leading to expansion in to less-favorable lands. Barriers to market are also causes of food insecurity in Africa (Mwanki, 2005; FAO, 2005). As he mentioned some barriers of market access were poor infrastructure, market standards, limited information, and requirements for large initial capital investments, limited product differentiation, and handicapping policies.

Diseases and infection are also identified as causes of food insecurity. Alex ( 2003) found that diseases such as malaria, tuberculosis and mainly HIV/AIDS not only reduce the man hours available to agriculture and household food acquisition, but also increase the burden of household in acquiring food.

Migration of male labor is also recognized as a cause of food insecurity. A study conducted in Lesotho village found that women and children suffered from lack of food and hygiene because women were too exhausted to cook and clean at times of peak agricultural work (Huss-Ashmore: cited in Driba, 1995 and Degafa, 2002). Haswell (1953) observed that growing cash crops at the expense of subsistence crops has largely contributed to seasonal food deficiency among the Gernieri in Gambia. He also observed that illness of adults at critical times in the production process adversely affects labor efficiency and productivity, which in turn contributes to seasonal food shortage.

Deterioration in the ecological conditions of production has also been seen as a cause of seasonal hunger in several African countries. Associated with this, Ogbu (1973) noted insufficient farm land, low yields on farms and high storage losses of staples as the principal causes of food shortage in Nigeria..

A similar research conducted by Toulmin (1996) noted that the people of Bambara of Kala in Mali face seasonal food shortages that are mainly induced by two principal factors: one of the factors is climatic, specifically low and highly variable rainfall making the people very vulnerable to crop failure. The second class of risk is demographic, consisting of high level of mortality, varying level of fertility and vulnerability of all producers to sickness and disability.

### **2.2.1.2. The Ethiopian Case**

Although investigations concerning farm households' food shortage have been limited, the situation in Ethiopia does not deviate much from the condition in other developing regions.

A combination of factors has resulted in serious and growing problem of food insecurity in Ethiopia. Adverse climate changes (drought) combined with human population pressure, environmental degradation, technological and institutional factors have led to a decline in the size of per capita land holding (MoFED, 2002). According to MoFED (2002) report the problem was exacerbated by policy induced stagnation of agriculture and internal conflict and instability in the past resulting into the widening of the food gap for more than two decades, which had to be bridged by food aid.

A research finding by Markos (1997) shows that "household's average cereal production during normal harvest year is persistently lower than annual food requirements and hence many households feed themselves from their farm outputs only for less than three-fourth of the year." Matha's (2002 :cited in Degafa, 2002) study in Meket, Habru and Gubalafto *Weredas* of North Wello found out that 30%, 21%, and 40% of the sample households, respectively, were unable to satisfy the food demand of their family for more than five months a year. Based on an empirical study in North Shewa, Yared (1997) argued that the seasonality of agriculture introduces fluctuations in the income, expenditure and nutritional patterns of peasant households.

Getachew (1995) concluded that households' risk of food insecurity and famines were greatly increased by long-term secular decline in resource endowment, combined with unfavorable food policy intervention. Emphasizing on subsistence farmers' food insecurity situation, he underlines that the prevailing inability of Ethiopia's small-scale agriculture to feed its population is mainly

generated by the neglect of the policy and the decline in access to productive resource upon which most of the livelihood are built.

In general, many of the natural and human induced factors that made Ethiopia a food-insecure country at the national level over the last few decades are mentioned in a paper by Degafa (2002) including fragile natural resource base, inadequate and variable rainfall, unimproved farming practices, inaccessibility to productive resources ( rural credit), diminishing land holdings and tenure insecurity, poor development of human resources, poor storage technology, inaccessibility to transport infrastructure, heavy work-load on women, poor health status, lower productivity of livestock, high level of unemployment, inappropriate use and non-integrated free food distribution of food aid, socio-cultural barriers, and lack of baseline information.

### **2.2.2. Determinants of household food security**

Those discussed in section 2.2.3 above are causes or determinants of food security at National, Regional or community levels. However, a study by Keshav (2006) shows that commonly used indicators of food security at the regional and national level or community level is often poor predictors of household food security. The study also made comparison among households based on depth and severity of food insecurity and found that socio-economic factors are the main determinants of food insecurity. The study concluded that both depth and severity of food insecurity are higher in occupational castes, small farms and less livestock holders, laborers, and households having minimum expense.

A number of studies made use of various methodologies to identify determinants of food security in different parts of Ethiopia. According to studies conducted by Shiferaw *et al.* (2003) and Webb *et al.* (1992); livestock ownership, farmland size, family labour, farm implements, employment opportunities, market access, level of technology application, level of education, health status, weather conditions, crop disease, rainfall, oxen ownership and family size were identified as major determinants of farm households' food security in Ethiopia.

A study by Haile *et al.* (2005) conducted in Koredegaga Peasant Association, Oromia Zone, identified that farmland size, per capita aggregate production, fertilizer application, household size,

ox ownership, and educational attainment of farm households heads had a significant influence on food security. The computed partial effects at sample means using results from the logistic regression model indicated that a unit change in farmers' access to fertilizer or educational level of household heads or farmer's access to land or access to family planning improve the probability of food security in the study area.

Another similar study by Ramarkishna *et al.* (2002) conducted in North Wollo revealed that per capita land holding, cereal production, livestock, educational level of household heads, fertilizer use and family size were the major determinants of food security. They constructed food balance sheet and food security causation was examined using a binary logistic regression model.

### **2.2.3 Generalizations of the Causes and Determinants of Food Insecurity**

From the theoretical and empirical causes and determinants of food insecurity, it can be generalized that food insecurity is a function of environmental crises, rapid population growth, poor assets basis, socio-cultural related issues, and poor access to market and infrastructure. Hence, in this sub-topic it is attempted to review relevant literatures particularly conducted in Ethiopia.

#### **2.2.3.1. Demographic Factors**

The population of Ethiopia is rising from time to time. Currently the Ethiopian population is about 74 million which grows by 2.6 % (CSA, 2008). According to CSA (2008) the average household size is also large when compared with other Sub-Saharan countries. At the micro level, *household size* is one of the factors expected to have influence on food security status of households. The majority of farm households in Ethiopia are small scale semi-subsistence producers with limited participation in non-agricultural activities since land holding size and financial capital to purchase agricultural inputs is very limited. Kidane (2005) in his work found that family size tends to exert more pressure on consumption than the labor it contributes to production.

Another demographic factor that strongly influences household food security is *sex* of the household head. Studies by Degafa (2002), Ramarkrisha *et al.* (2002) and Kidane *et al.* (2005)

independently conducted in different parts of rural Ethiopia came out with common conclusion that the livelihood of female headed households was disadvantaged when compared with their male counterparts. This is due to the fact that, the researchers justify, female household heads have limited access to livelihood assets like land, education, saving, labor force and oxen (drought power), livestock and credit services.

### **2.2.3.2. Environmental Crises**

The combined effect of land based resources degradation like deforestation, soil erosion, flooding, and loss of agricultural and pasture land leads to production decline (Getachew, 1995). Rapid population growth and recurrent drought are causing serious resource degradation. Markos (1997) and Fitsum *et al.* (2002) described that the seriousness of shortage of productive ( fertile) land in the highland areas, coupled with population pressure, have forced the cultivation of the steep and moderate slopes which are highly degraded because of soil erosion.

Climate is one of the important elements of the natural environment that positively or negatively affects the food security status of rural households. Many studies indicated that inadequate and erratic rainfall is one of the environmental phenomena, causing food crises in many rain fed farming and drought prone areas across the world. In Ethiopia more than 95% of food grain production is from rain fed subsistence farm (Osman, 2005: cited in Adane, 2008). A study conducted in Ethiopia by Devereux (2002) revealed that a 10% decline in rainfall below its long-term average reduces national food production by 4.4%.

### **2.2.3.3. Poor Asset Base of the Rural Households**

In countries like Ethiopia where agricultural sector employed 89% of the labor force and contributed 56% of GDP and 67 % of export earning (Devereux, 2000), land is an indispensable resource. Given the level of agricultural technology, certain minimum land holding size is required to produce sufficient production. Yared (1999) in his study in Wagda concluded that household land holdings play the most fundamental role in determining grain and animal production in the rural economy. He added that in Wagda, access to drought power and labor participation are influenced by the size of the land people owned.

Farm equipments and basic infrastructure are among the physical capitals that influence the day to day activities of rural households as producers and consumers. Dulla (2007) stated that ownership of machinery and equipment enables households to raise labor and land productivity and is especially helpful for households with relatively high opportunity costs for labor, such as those pursuing off-farm employments.

Fertilizer use is used by most studies as a proxy for technology. Literatures on roles of fertilizer in agricultural productivity found that fertilization of farmland can boost agricultural production and influence the food security status of a household. Study by Kidane *et al.* (2005) concluded that the shift from non-fertilizer user to fertilizer user increased the probability of food security from 33.8% to 44.3 %, but in the country those who apply fertilizer are insignificant due to their limited purchasing power.

#### **2.2.3.4. Socio-Cultural Factors**

Education has a tremendous influence on the food security status of households. Educational attainment by the household head could lead to awareness of the possible advantages of modernizing agriculture by means of technological inputs; enable them to read instructions on fertilizer packs and diversification of household incomes which, in turn, would enhance household's food supply ( Kidane *et al.*, 2005).

Socio-cultural events such as eating habit and food preference, cultural ceremonies and festivals also influence the food security status of the given communities and way of saving or expenditure, also directly or indirectly affects the food security situation of that particular community.

#### **2.2.3.5. War or Conflict**

Different literatures revealed that the present day famine in Africa are largely the result of military conflict that arises due to oppressive, unaccountable, and non participatory governments. The experience of Sudan, Liberia, Ethiopia, Chad, Rwanda, Burundi, and Somalia depicted how war disrupts the normal functioning of the economy, social and political situations (Salih, 1994; Fasil, 2005). According to Getachew (1995), in Ethiopia and in Sudan alone 10 million people were affected by the civil war and estimates prevailed that more than 5 million people died.

Regarding resource misallocation, before 1991, in Ethiopia more than 50% of GDP was spent on the war effort while food security and other economic and social development agendas were neglected (MoFED, 2000; Adane, 2008).

#### **2.2.3.6. Access to Infrastructure**

Access to infrastructure such as market center and roads promote livelihood diversification and agriculture intensification. Adequate infrastructure, especially main and feeder roads that improve access to necessary input-fertilizer, seed, pesticide chemicals and other agricultural implements are very indispensable (Osman, 2003). Although, the current government has made a significant progress particularly in road development, the sector is still weak even compared with the African average. World Bank (2007) reported that due to lack of proper and on time transportation facilities post harvest total production loss reached up to 30%.

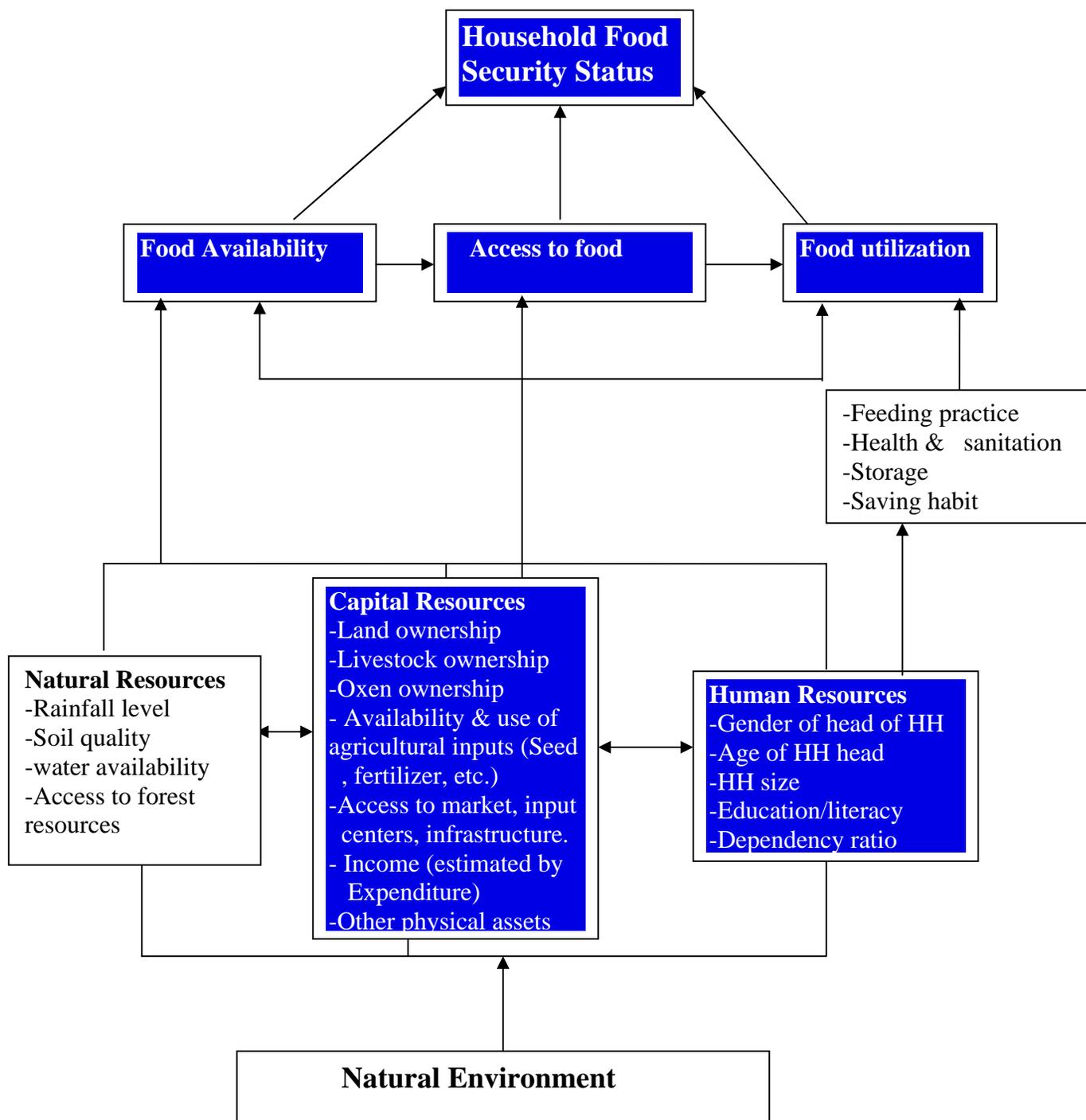
Generally, as indicated in many literatures, inadequate infrastructures and social services development such as road, transportation, communication, electrification, education and health services and agricultural services would be major challenges to sustain the growth of agricultural production and food security.

### **2.3. Comments on the Reviewed Literatures**

Much of the reviewed literature on household food security concentrated on describing qualitatively and quantitatively the extent of household food security and identifying the factors and examining their implications. Almost all reviewed studies applied logistic regression in modeling relationships between variables. However, the central task of regression analysis: the parameter estimation techniques and variable selection methods were not addressed. Most of the reviewed model did not check model adequacy: detection and treatment of outliers, influence diagnostics and multicollinearity.

Almost all reviewed studies did not examine the effect of factors in discriminating food secure households from food insecure households. Hence, in this study in addition to the prediction, a model that can discriminate and classify households based on the discriminant factors will be used taking into account the limitations described in the reviewed literature.

Figure 2.1: Food Security status Framework



Adopted from Frank et al. (1999) and Majda (2000) with major modification

## **CAPTER THREE: METHODOLOGY**

### **3.1 Data Source**

The data for this study was taken from Household Income, Consumption and Expenditure (HICE) survey and Welfare Monitoring (WM) survey which were conducted by Central Statistical Agency (CSA) in 2004/5.

For both surveys, the list of households obtained from the 2001/2 Ethiopian Agricultural Sample Enumeration (EASE) was used as a frame to select Enumeration Areas (EAs) from the rural part of the country and the 2004 Ethiopian Urban Economic Establishment Census (EUEEC) was used as a frame in order to select sample EAs from the urban part of the country. A fresh list of households from each rural and urban EAs was prepared at the beginning of the survey period. This list was, thus, used as a frame in order to select households from sample EAs.

For the purpose of the survey the country was divided into three broad categories. That is; rural, major urban centers and other urban center categories. The first category consists of the rural areas of eight regional states and two administrative councils (Addis Ababa and Dire Dawa) of the country except Gambella region. Each region was considered to be a domain (Reporting level) for which major findings of the survey are reported. This category comprises 10 reporting levels. A stratified two-stage cluster sample design was used to select samples in which the primary sampling units (PSUs) were EAs. Twelve households per sample EA were selected as a second stage sampling unit (SSU) to which the survey questionnaire were administered. The second category includes all regional capitals (except Gambella region) and four other urban centers that have relatively larger population sizes. Each urban center in this category was considered as a reporting level. A stratified two-stage cluster sample design was also adopted in this instance. The primary sampling units were EAs of each urban center. Sixteen households from each sample EA were finally selected as a secondary sampling unit. The third category includes other urban centers in the country other than those in the second category. Unlike the above two categories a stratified three-stage cluster samples design was adopted to select samples from this category.

Totally 2,016 EAs and 24,192 households were selected for the WM survey including 797 EAs and 9,564 households for the HICE survey from the first category. Sample EAs of each reporting level, in this category, were selected using probability proportional to size (PPS) with systematic sampling techniques, size being number of households from the 2001/2 EASE. Twelve households per EA were systematically selected from the fresh list of households prepared at the beginning of the survey.

The rural part of Tigray region is one of the reporting levels of the first category. Hence, from rural Tigray 164 EAs and 1968 households were selected for WM survey out of which 71 EAs and 852 households (HHs) were selected for HICE survey. This study used the 852 rural HHs which are common in both surveys. These 852 rural HHs were selected from all zones of the region except from Mekelle Zone. However, out of the 852 selected households, information was not gathered on 24 households due to different reasons. Since the non-respondent households were found randomly in all zones of the region and since the response rate is above 97%, this study excluded the non-respondent households. Hence, the sample size of this study is 828 households (that is,  $n=828$ ).

Both surveys provide information regarding the basic population characteristics such as sex, age, household size, marital status, education and employment. In particular, the HICE survey includes information regarding the household consumption expenditure: food and non-food expenditure, as well as quantities consumed, payments and income. The food consumption expenditure is calculated by adding up the value of subsistence food consumption to cash expenditure on food. On the other hand, the WM survey provides information that mainly shows the status of education, health, and vulnerability i.e., access to health, education and other facilities. Ownership of physical assets, facility of agricultural extensions, access to market and, in general, the living standard of households etc. are also included in the WM survey.

Moreover, the HICE survey data was collected in two rounds of the year. That is the slack and peak period to represent the living condition of the family in the whole year (CSA, 2004/5).

## 3.2 Variables in the study

The dependent and independent variables that were considered to affect the status of household food security were selected based on experiences from the available similar studies and the available data on the subject.

### 3.2.1 The Dependent variable

The dependent variable in this study is Household Food Security (HFS) status. Consumption-based rather than income-based measure of HFS status is used in this study. This is because consumption better captures long-run welfare, and it better reflects household's ability to meet their basic needs. Consumption is preferable to measure HFS than income because it is less vulnerable to seasonality and life-cycle, less vulnerable to measurement errors because respondents have less reasons to lie, it is closer to the utility that people effectively extract from income, and for the poor most of income is consumed.(CSA, 2005; FAO, 2002).

The HFS status was determined using the consumption approach based on the 2004/5 HICE survey conducted by CSA. Following this approach, household food security status was set on the basis of the caloric content of consumed food items. To do this, first the bundle of food items consumed by households was listed and measured in terms of 100 gram solid food using conversion factors for the liquid and semi-liquid food items. Second, for each food item a caloric content value was assigned based on the 1998 food composition table by Ethiopian Nutrition and Health Research Institute (ENHRI) which is given in Table A3 (ANNEX A). Total Net Calorie (TNC) was estimated based on the total edible portions of weights of consumed food items for each household. Third, due to differences in household compositions in terms of age and sex, there was a need to adjust the household size to adult equivalent household size. Adult equivalence was developed by World Health Organization (WHO) considering the nutritional requirements of an individual by age and gender. Adult equivalence table given in Table A2 (ANNEX A) is used as a reference to calculate adult equivalent household size in this study.

Finally, the HFS Status was defined based on the consumption per adult equivalent per day. This is given as:

$$HFS_i = \frac{TNC \text{ consumed by household in one day}}{\text{Adult equivalent household size}} \quad \text{where } i = 1, 2, \dots, 828 \quad (2.1)$$

Following the Food Security Strategy of the Federal Democratic Republic of Ethiopia (1996) and FAO (2002), 2100 Kcal per day was assumed to be the minimum energy requirement enabling an adult to lead a healthy and moderately active life in Ethiopia, particularly in the study area.

Households whose consumed calories were found to be greater than their calorie requirement were regarded as food secure and assigned a value of 1, while households who faced with calorie deficiency during the study year were regarded as food insecure and they were assigned a value of 0. Hence, the dependent variable, food security status of the  $i^{th}$  household, was measured as a dichotomous variable:

$$Y_i = \begin{cases} 1, & HFS_i \geq 2100 \text{ Kcal (Food secure)} \\ 0, & HFS_i < 2100 \text{ Kcal (Food insecure)} \end{cases} \quad (2.2)$$

where  $Y_i$  is food security status of the  $i^{th}$  household,  $i = 1, 2, \dots, 828$

### 3.2.1.1 Measurement of food insecurity

Consumption below the minimum level of calorie requirement indicates food insecurity. The food insecurity measures used in this study, adopted from FAO (2002) and Keshav (2006) are head count ratio, food insecurity gap, and squared food insecurity gap to capture successively more detailed aspect of the food insecurity at the household level.

Head count ratio is given as  $IFI = m/n \times 100$  where,  $IFI$  is Incidence of Food Insecurity,  $m =$  Number of Food Insecure Households, and  $n =$  Total Number of Households under study.

Food Insecurity Gap of  $i^{th}$  food insecure household ( $FIG_i$ ) is defined as:

$$FIG_i = \frac{TCR_i - TCC_i}{TCR_i} \quad \text{where, } TCR_i \text{ Total Calorie Requirement for } i^{th} \text{ food insecure household}$$

and  $TCC_i$  denotes the Total Calorie Consumption by  $i^{th}$  food insecure household. Total Food

Insecurity Gap (TFIG), which indicates the depth of food insecurity among the food insecure households, is expressed as  $TFIG = \sum_{i=1}^m \frac{FIG_i}{m}$ . Finally the Squared Food Insecurity Gap (SFIG),

which indicates severity of food insecurity among the food insecure households, is given as:

$$SFIG = \sum_{i=1}^m \frac{(FIG_i)^2}{m}.$$

### 3.2.2 Explanatory Variables

Based on the reviewed literatures, some of the common predictors that are expected to influence rural household's food security in the study area could be categorized into Demographic and Socio-Economic variables.

Table 3.1: Demographic variables

| Variables      | Description                 | Values                 |
|----------------|-----------------------------|------------------------|
| AGE ( $X_1$ )  | Age of head of household    | Years                  |
| GEND ( $X_2$ ) | Gender of head of household | 0 = Female<br>1 = Male |
| HHSZ ( $X_3$ ) | Household size              | Number                 |
| CDR ( $X_4$ )  | Child dependency ratio      | Number                 |

Table 3.2: Socio-Economic Variables

| Variables        | Description                                 | values  |
|------------------|---|---------|
| EDUC ( $X_5$ )   | Educational level of head of household      | Years   |
| EXPEND ( $X_6$ ) | Average monthly expenditure of a household  | Birr    |
| FLSZ ( $X_7$ )   | Farm land size of a household               | Hectare |
| TLU ( $X_8$ )    | Total number of Livestock ( Excluding oxen) | TLU     |
| OXEN ( $X_9$ )   | Number of oxen owned by the farm household  | Number  |

| Variables           | Description                              | Values            |
|---------------------|--|-------------------|
| FRTLZR ( $X_{10}$ ) | Use of Fertilizer                        | 0= No<br>1= Yes   |
| AGREXT ( $X_{11}$ ) | Access to agricultural extension service | 0 = No<br>1 = Yes |
| SEED ( $X_{12}$ )   | Use of improved Seeds                    | 0 = No<br>1 = Yes |
| DIST ( $X_{13}$ )   | Distance to input source                 | Kilometer         |

### 3.2.2.1. Description of Explanatory Variables

**Age of head of household (AGE):** Age is measured in years. Older people have relatively richer experiences of the social and physical environments as well as greater experience of farming activities (Haile *et al.*, 2005). That is, when heads get higher age, they are expected to have stable economy in farming. Moreover, older household heads are expected to have better access to land than younger heads, because younger men either have to wait for land redistribution, or have to share land with their families. However, Babatunde (2007) and other related studies stated that young head of households were stronger and were expected to cultivate larger-size farm than old heads. Hence, the expected effect of age on HFS could be positive or negative.

**Gender of head of household (GEND):** households headed by female, according to the reviewed literatures, have higher probability of being food insecure. Hence, in this study gender is expected to be positively related with HFS.

**Household size (HHSZ):** increasing family size, according to reviewed literatures, tends to exert more pressure on consumption than the labour it contributes to production. Thus a negative correlation between household size and food security is expected in this study.

**Child dependency ratio (CDR):** It is defined as the ratio of number of persons less than 15 years of age to the number of persons in the age group 16 - 64 years (labour force). In subsistence farming, households with larger labour supplies are better positioned to increase productivity of their land (Hofferth, 2003: cited in Haile *et al.*, 2005). Availability of a relatively larger labour

force, regardless of farmland size, can be an advantage to those households who strive to achieve food security. This is because the labour force can participate in non-agricultural activities that enhance their income. However, the larger CDR the household has the more food insecure the household is expected to be. Hence, child dependency ratio is expected to have a negative impact on HFS.

**Educational level of head of household (EDUC):** Education is a social capital, which could impact positively on household ability to take good and well-informed production and nutritional status (Babatunde *et al.*, 2007). Based on Amaza *et al.* (2006) and other literatures, the higher the educational level of household head, the more food secure the household is expected to be.

**Average monthly expenditure of household (EXPEND):** The average monthly expenditure of a household, measured in Birr, is defined as the cost paid for food and non-food consumption, for agricultural inputs such as for fertilizers, improved seed, pesticides, insecticides and other expenses by the household in one month. A household with higher income spends more. Hence, household expenditure is expected to be positively correlated with household food security.

**Farm land size of a household (FLSZ):** Farm land size is the total farm land cultivated by the household measured in hectares. According to Haile *et al.* (2005) and Babatunde *et al.* (2007) and other literatures, food production can be increased extensively through expansion of areas under cultivation. It is thus expected that households with larger farm size are more likely to be food secure than those with smaller farm size. The expected effect of farm land size on HFS is positive.

**Livestock ownership:** It is measured by the number of Tropical Livestock Unit (TLU). Livestock contribute to household's economy in different ways, e.g. as a source of pulling power, source of cash income, source of supplementary food, and means of transport. Besides, livestock are considered as a means of security and means of coping during crop failure and other calamities (Haile *et al.*, 2005). Livestock provides not only food for the producers, but also a large number of other products which could be sold or consumed by the livestock owner to provide nutrition, income, traction and fuel. The major products of livestock include pulling power, meat, milk, eggs, manure which is used as fertilizer or fuel, feather, fiber, hides, and horns. Households who own

livestock have good food security status as well as sustainable farming (Ramakrishna and Assefa, 2002). Livestock ownership was measured by the number of TLU owned by the household during the study time excluding oxen. Conversion factors were used in order to change each livestock of a household to its equivalent TLU, which are given in Table A1 (ANNEX A). Positive correlation is expected between livestock ownership and HFS.

**Number of oxen ownership (OXEN):** Number of ploughing oxen is another determinant of the food security status of households. Oxen serve as a source of traction in many developing countries, thereby significantly affecting household's crop production. Animal traction power enables households to cultivate greater source of land and to execute agricultural operations timely (Hail *et al.*, 2005). Therefore, a positive relationship between oxen ownership and food security is expected in this study.

**Use of Fertilizer (FRTLZR):** According to literatures, fertilization of farm land could increase agricultural production and influence positively the food security status of a household.

**Access to agricultural extension service (EXTN):** According to the reviewed literatures, households that have access to extension service were to be more food secure than those who did not have access. Amaza *et al.* (2006) stated that access to extension service enhances households' access to better crop production techniques, improved input as well as other production incentives and these positively affect their food security status. Hence, household's access to extension service is expected positively related with food security.

**Use of improved seed (SEED):** Improved genetic resource of seeds is essential to increase agricultural production. A high quality of seeds of improved or indigenous crops adjusting with the ecological and environmental conditions boosts the over all crop production. Use of improved seed is expected to have a positive effect on HFS.

**Distance to input source (DIST):** Agricultural inputs such as fertilizers, seeds, pesticides, chemicals, and other agricultural implements are very important for farmers to produce more crops and breed animals. However, if the distance to the source of these inputs is very long and the

infrastructure is very poor, farm households may not access these inputs timely. Hence, distance to input source is expected to have negative impact on HFS.

### 3.3 Model Selection

A limitation of ordinary linear models is the requirement that the dependent variable is numerical rather than categorical. But many interesting variables are categorical—patients may live or die, people may pass or fail exams, households may be food secure or insecure and so on. A range of techniques have been developed for analyzing data with categorical dependent variables, including discriminant analysis, probit analysis, log-linear regression and logistic regression.

Linear Discriminant Analysis (LDA) and Logistic Regression (LR) are widely used multivariate statistical methods for analysis of data with categorical outcome variables. Both of them are appropriate for the development of linear classification models, i.e. models associated with linear boundaries between the groups (Tabachnick and Fidell, 2007).

Nevertheless, the two methods differ in their idea. While LR makes no assumption on the distribution of the explanatory data, LDA has been developed for normally distributed explanatory variables. It is therefore reasonable to expect LDA to give better results when the normality assumptions are fulfilled, but in all other assumptions LR should be more appropriate (Maja *et al.*, 2004).

Linear discriminant analysis can be used to determine which variable discriminates between two or more classes, and to derive a classification model for predicting the group membership of new observations. For each of the groups, LDA assumes that the explanatory variables are normally distributed with equal covariance matrices. The simplest LDA has two groups. To discriminate between them, a linear discriminant function that passes through the centroids of the two groups can be used.

When the covariates are sampled from the normal distribution, LDA of course seems to be the more appropriate method (Maja *et al.*, 2004). This assumption is actually met when the sample size is large. This is because, as the sample size increases, the sampling distributions become more

stable which leads to better results for the LDA. In this case, LR can also give results which are closer to the results of LDA.

Logistic regression can handle both categorical and continuous predictor variables, and the predictors do not have to be normally distributed, linearly related, or of equal variance within each group. Discriminant analysis requires continuous predictors. However, categorical predictor variables are also possible to use in LDA (Tabachnick and Fidell, 1996).

As most of the predictor variables in this study are continuous variables and since the sample size is large, the assumption of normality is expected to hold. Hence, the expected appropriate models for this study could be both logistic regression and discriminant analysis. However, under the assumption of normality, discriminant analysis was expected to reveal better results than logistic regression. Hence, this study uses discriminant analysis to meet the objectives set in this study.

### **3.4 Discriminant Analysis**

Discriminant Analysis (DA) is a dimension reduction method that can be used to identify a linear combination of variables that produces the greatest distance between groups. DA is usually used to predict group membership in naturally occurring groups. It answers the questions: Can a combination of variables be used to predict group membership? Usually several variables are included in a study to see which ones contribute to the discrimination between groups. DA is conceptually similar to logistic regression for multivariate data, and it is computationally similar to Multivariate Analysis of Variance (MANOVA).

Discriminant analysis and classification are multivariate techniques concerned with separating distinct sets of objects and with allocating new objects to previously defined groups. The goal of discrimination is to describe graphically or algebraically, the differential features of objects from several known collections (populations) and the goal of classification is to sort objects into two or more labeled classes.

### 3.4.1. Assumptions of Discriminant Analysis

- ❖ **Multivariate Normality-** DA assumes that the various independent variables in each group and all linear combinations of them are normally distributed. However, violations of the normality assumption are not “fatal” and the resultant significance test is reliable as long as non-normality is caused by skewness and not outliers. In univariate tests robustness against violations of the assumption is assured when the degree of freedom for error is 20 or more for equal group sizes. Moreover, if there are at least 20 cases in the smallest group the test is robust to violations of multivariate normality even when there are unequal group sizes (Tabachnick and Fidell, 1996).
- ❖ **Homogeneity of Covariance Matrices-** DA assumes that the variance/covariance matrix in each group of the design is sampled from the same population so they can be reasonably pooled together to make an error term. When inference is the goal, discriminant analysis is robust to violations of this assumption (Tabachnick and Fidell, 1996). When classification is the goal, then the analysis is highly influenced by violations because subjects will tend to be classified in to groups with the largest dispersion. If this assumption is violated, appropriate remedies include transforming the data, using separate matrices during classification, using quadratic discriminant or using non-parametric approaches to classification.

### 3.4.2 Theoretical Development of Discrimination and Classification

Classification analysis is concerned with the development of rules for allocating or assigning observations into one or more groups. A classification rule usually requires more knowledge about the parametric structure of the groups. The goal of classification analysis is to create rules for assigning observations to groups that minimize the total probability of misclassification or the average cost of misclassification.

Because linear discriminant functions are often used to develop classification rules, the goals of the two processes tend to overlap and some authors use the term classification analysis instead of

discriminant analysis. Because of the close association between the two procedures we treat them together in this sub-section.

### 3.4.2.1 Separation and classification for two populations

It is convenient to label the two populations (classes) as  $\pi_1$  and  $\pi_2$ . The objects are ordinarily separated or classified on the basis of measurements on, for instance,  $p$  associated random variables  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ . The observed values of  $\mathbf{X}$  differ to some extent from the one class to the other. We can think of the totality of values from the first class as being the population of  $\mathbf{x}$  values for  $\pi_1$  and those from the second class as the population of  $\mathbf{x}$  values for  $\pi_2$ . These two populations can then be described by probability density functions  $f_1(\mathbf{x})$  and  $f_2(\mathbf{x})$ , and, consequently, we can talk of assigning observations to populations or objects to classes interchangeably.

For allocating an observation (household) to one of the two populations  $\pi_1$  and  $\pi_2$ , we need a set of measurements that has been made on it. According to Krazanowski (1980), a large number of individuals tend to be classified into their respective groups, and hence the classification rule should be chosen in such a way as to minimize the expected consequence of the mistakes made in this series of allocations. Suppose the set of measurement (data matrix)  $n$  units is partitioned in to two groups: one comprising  $n_1$  units, constituting a random sample from a P-variate population  $\pi_1$ , while the second group, comprising  $n_2$  units, constituting a random sample from a P-variate population  $\pi_2$ .

Let the cost of an observation from  $\pi_i (i = 1, 2)$  is misclassified as  $\pi_j (j = 1, 2)$  be  $C(j|i)$ . The cost of such correct classification  $C(i|i)$ , is zero and the cost of incorrect classification  $C(j|i)$  is greater than zero for  $i \neq j$ . Various classification procedures have been proposed to determine optimum allocation rules. One of such procedures is to minimize the cost of assignment to the groups.

Let the probability density functions associated with the  $P \times 1$  vector random variable  $\mathbf{X}$  for the populations  $\pi_1$  and  $\pi_2$  be  $f_1(\mathbf{x})$  and  $f_2(\mathbf{x})$ , respectively. An object, with associated measurements  $\mathbf{x}$ , must be assigned to either  $\pi_1$  or  $\pi_2$ . Let  $\Omega$  be the sample space; that is, the collection of all possible observations  $\mathbf{x}$ . Let  $R_1$  be the set of  $\mathbf{x}$  values for which we classify objects as  $\pi_1$  and  $R_2 = \Omega - R_1$  be the remaining  $\mathbf{x}$  values for which we classify objects as  $\pi_2$ . Since every object must be assigned to one and only one of the two populations, the sets  $R_1$  and  $R_2$  are mutually exclusive and exhaustive. Let  $P(j|i)$  be the conditional probability of classifying an object as  $\pi_j$  when, in fact, it is from  $\pi_i$  (for  $i, j = 1, 2$ ). Hence, the conditional probability of classifying an object as  $\pi_2$  when, in fact, it is from  $\pi_1$  is

$$P(2|1) = P(\mathbf{X} \in R_2 | \pi_1) = \int_{R_2} f_1(\mathbf{x}) d\mathbf{x} \quad (3.1)$$

Similarly, the conditional probability of classifying an object as  $\pi_1$  when it is really from  $\pi_2$  is

$$P(1|2) = P(\mathbf{X} \in R_1 | \pi_2) = \int_{R_1} f_2(\mathbf{x}) d\mathbf{x} \quad (3.2)$$

Let  $P_1$  be the prior probability of  $\pi_1$  and  $P_2$  be the prior probability of  $\pi_2$ , where  $P_1 + P_2 = 1$ . The overall probabilities of correctly or incorrectly classifying objects can be derived as the product of the prior and conditional classification probabilities. That is

$$\begin{aligned} P(\text{correctly classified as } \pi_1) &= P(\text{observation comes from } \pi_1 \text{ and} \\ &\quad \text{is correctly classified as } \pi_1) \\ &= P(\mathbf{X} \in R_1 | \pi_1) P(\pi_1) = P(1|1) P_1 \end{aligned}$$

$$\begin{aligned} P(\text{misclassified as } \pi_1) &= P(\text{observation comes from } \pi_2 \text{ and} \\ &\quad \text{is misclassified as } \pi_1) \\ &= P(\mathbf{X} \in R_1 | \pi_2) P(\pi_2) = P(1|2) P_2 \end{aligned}$$

$$\begin{aligned} P(\text{correctly classified as } \pi_2) &= P(\text{observation comes from } \pi_2 \text{ and} \\ &\quad \text{is correctly classified as } \pi_2) \\ &= P(\mathbf{X} \in R_2 | \pi_2) P(\pi_2) = P(2|2) P_2 \end{aligned}$$

$$\begin{aligned}
P(\text{misclassified as } \pi_2) &= P(\text{observation comes from } \pi_1 \text{ and} \\
&\text{is misclassified as } \pi_2) \\
&= P(\mathbf{X} \in R_2 \mid \pi_1) P(\pi_1) = P(2 \mid 1) P_1
\end{aligned}
\tag{3.3}$$

Classification schemes are often evaluated in terms of their misclassification probabilities, but this ignores misclassification cost. A rule that ignores costs may cause classification problems. The costs of misclassification can be defined by a cost matrix.

|                 |         | Classify as |          |
|-----------------|---------|-------------|----------|
|                 |         | $\pi_1$     | $\pi_2$  |
| True population | $\pi_1$ | 0           | C(2   1) |
|                 | $\pi_2$ | C(1   2)    | 0        |

(3.4)

The costs are : zero for correct classification, C(1 | 2) when an observation from  $\pi_2$  is incorrectly classified as  $\pi_1$ , and C(2 | 1) when a  $\pi_1$  observation is incorrectly classified as  $\pi_2$ . For any rule, the average, or *expected cost of misclassification* (ECM) is provided by multiplying the off-diagonal entries in (3.4) by their probabilities of occurrence obtained from (3.3). Consequently,

$$\text{ECM} = C(2 \mid 1) P(2 \mid 1) P_1 + C(1 \mid 2) P(1 \mid 2) P_2
\tag{3.5}$$

A reasonable classification rule should have an ECM as small as possible.

### 3.4.2.2 Optimal Classification rules for two populations

In the preceding section, it was suggested that a sensible classification rule could be determined by minimizing the ECM. In other words, the assignment regions  $R_1$  and  $R_2$  must be chosen so that the ECM is as small as possible. The regions  $R_1$  and  $R_2$  that minimize the ECM are defined by the values  $\mathbf{x}$  for which the following inequalities hold.

$$R_1 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \left[ \frac{C(1|2)}{C(2|1)} \right] \left[ \frac{P_2}{P_1} \right]$$

[Density ratio]  $\geq$  [Cost ratio] [Prior probability ratio]

$$R_2 : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \left[ \frac{C(1|2)}{C(2|1)} \right] \left[ \frac{P_2}{P_1} \right]$$

[Density ratio] < [Cost ratio] [Prior probability] (3.6)

When the prior probabilities are unknown, they are often taken to be equal and the minimum ECM rule involves comparing the ratio of the population densities to the ratio of the appropriate misclassification costs. If the misclassification cost ratio is indeterminate, it is usually taken to be unity and the population density ratio is compared with the ratio of the prior probabilities. The optimal classification regions are determined simply by comparing the values of the density functions. In this case, if  $\mathbf{x}_0$  is a new observation and  $f_1(\mathbf{x}_0) / f_2(\mathbf{x}_0) \geq 1$  [that is,  $f_1(\mathbf{x}_0) \geq f_2(\mathbf{x}_0)$ ], we assign  $\mathbf{x}_0$  to  $\pi_1$ . On the other hand, if  $f_1(\mathbf{x}_0) / f_2(\mathbf{x}_0) < 1$  [ $f_1(\mathbf{x}_0) < f_2(\mathbf{x}_0)$ ], we assign  $\mathbf{x}_0$  to  $\pi_2$ . It is equivalent to assuming equal prior probabilities and equal misclassification costs for the minimum ECM rule (Johnson, 2002).

Criteria other than the expected cost of misclassification can be used to derive optimal classification procedures. For example, one might ignore the costs of misclassification and choose  $R_1$  and  $R_2$  to minimize the *total probability of misclassification* (TPM),

$$\begin{aligned} \text{TPM} &= P(\text{misclassifying a } \pi_1 \text{ observation or misclassifying a } \pi_2 \text{ observation}) \\ &= P(\text{observation comes from } \pi_1 \text{ and is misclassified}) \\ &\quad + P(\text{observation comes from } \pi_2 \text{ and is misclassified}) \end{aligned}$$

This problem is mathematically equivalent to minimizing the expected cost of misclassification when the costs of misclassification are equal.

### 3.4.2.3 Classification with two multivariate normal populations

We now assume  $f_1(\mathbf{x})$  and  $f_2(\mathbf{x})$  are multivariate normal densities; the first with mean vector  $\boldsymbol{\mu}_1$  and covariance matrix  $\boldsymbol{\Sigma}_1$  and the second with mean vector  $\boldsymbol{\mu}_2$  and covariance matrix  $\boldsymbol{\Sigma}_2$ . Assuming  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ , Fisher's linear discriminant function can be used for classification since it was developed under the assumption that the two populations, whatever their form, have a

common covariance matrix. Let us also assume that the population parameters  $\boldsymbol{\mu}_1$ ,  $\boldsymbol{\mu}_2$  and  $\boldsymbol{\Sigma}$  be all known and the joint densities of

$$\mathbf{X}' = [X_1, X_2, \dots, X_p] \quad \text{for populations } \pi_1 \quad \text{and} \quad \pi_2 \quad \text{be given by}$$

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right] \quad \text{for } i = 1, 2 \quad (3.7)$$

Then, the density ratio is

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = \frac{\exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_1)\right]}{\exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2)\right]}$$

$$= \exp\left\{\frac{-1}{2}\left[(\mathbf{x} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_1) - (\mathbf{x} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2)\right]\right\} \quad (3.8)$$

Hence, the regions in (3.6) that minimize ECM become

$$R_1: \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2) + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2)\right] \geq \left[\frac{C(1|2)}{C(2|1)}\right] \left[\frac{P_2}{P_1}\right]$$

$$R_2: \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2) + \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_2)\right] < \left[\frac{C(1|2)}{C(2|1)}\right] \left[\frac{P_2}{P_1}\right]$$

After taking natural logarithm of equation (3.8) and rearranging the vectors, the optimum allocation rule is given as follows:

Allocate  $\mathbf{x}_0$  to  $\pi_1$  if

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{x}_0 - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq \ln \left[\frac{C(1|2)}{C(2|1)}\right] \left[\frac{P_2}{P_1}\right] \quad (3.9)$$

Otherwise allocate  $\mathbf{x}_0$  to  $\pi_2$ .

By denoting the expression in the left hand side of equation (3.9) as:

$$L(\mathbf{x}) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1} \mathbf{x} - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2), \quad \text{this expression is a linear function of}$$

the observation vector ( $\mathbf{x}$ ) and hence it is known as Linear Discriminant Function (LDF). In fact,

$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}$  is a row vector,  $\boldsymbol{\alpha}' = (\alpha_1, \alpha_2, \dots, \alpha_p)$ . If  $\mathbf{x}_0$  consists of  $(x_{01}, x_{02}, \dots, x_{0p})$ , then the linear discriminant function is given as:

$L(\mathbf{x}_0) = \alpha_0 + \alpha_1 x_{01} + \alpha_2 x_{02} + \dots + \alpha_p x_{0p}$ , where  $\alpha_0$  is a constant and is given by

$$\alpha_0 = -\frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2).$$

IF the assumption of equal population covariance was violated (i.e., when  $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$ ) the function would be a quadratic discriminant function (see the details in Johnson, 2002). In such conditions, quadratic discriminant analysis controls the variability in each group and provides reliable results. Alternatively, it is appropriate to use logistic regression rather than discriminant analysis.

### 3.4.3 Applications of Discriminant and Classification Analysis

From literature and theoretical concepts, discriminant analysis was found to be an appropriate method of analysis in studying the effect of factors on the status of household food security. The major methods that will be used in the analysis are discussed in this section. However, before dealing with the methods it is better to introduce some concepts and notations that are used in the analysis.

As was used in the theoretical part,  $\pi_1$  and  $\pi_2$  refer to the first and second groups, respectively. In this study, the term *group* is used to represent either a population or sample from the population. For  $n$  observations and  $p$  associated vector random variables  $X_1, X_2, \dots, X_p$ , the  $n \times P$  matrix of explanatory variables is given as

$$\mathbf{X}_{n \times P} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]'$$

In multivariate analysis  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  usually form a random sample,

where  $\mathbf{x}'_i = [x_{i1}, x_{i2}, \dots, x_{ip}]$ ,  $x_{ij}$  is the value of the  $j^{th}$  variable on the  $i^{th}$  observation.

Let  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$  denote the population mean vectors for  $\pi_1$  and  $\pi_2$ , respectively. Similarly, the covariance matrices are denoted as  $\boldsymbol{\Sigma}_1$  for group one and  $\boldsymbol{\Sigma}_2$  for group two. Under the assumption of equal covariance, the common covariance matrix can be given as  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ .

In most practical situations, the population parameters,  $\boldsymbol{\mu}_1$ ,  $\boldsymbol{\mu}_2$  and  $\boldsymbol{\Sigma}$  are rarely known.

Wald (1944) and Anderson (1984) have suggested estimating the population parameters by their sample counterparts. Suppose that we have  $n_1$  observation of the multivariate random variable  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$  from  $\pi_1$  and  $n_2$  measurements of their quantity from  $\pi_2$ , where  $n = n_1 + n_2$ . The respective data matrices are

$$\mathbf{X}_{1(n_1 \times p)} = \begin{bmatrix} \mathbf{x}'_{11} \\ \mathbf{x}'_{12} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}'_{1n_1} \end{bmatrix} \quad \text{and} \quad \mathbf{X}_{2(n_2 \times p)} = \begin{bmatrix} \mathbf{x}'_{21} \\ \mathbf{x}'_{22} \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}'_{2n_2} \end{bmatrix} \quad (3.10)$$

From these data matrices, the sample mean vectors and covariance matrices are determined by

$$\begin{aligned} \bar{\mathbf{x}}_{1(p \times 1)} &= \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{x}_{1i}; \quad \mathbf{s}_{1(p \times p)} = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)(\mathbf{x}_{1i} - \bar{\mathbf{x}}_1)' \\ \bar{\mathbf{x}}_{2(p \times 1)} &= \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{x}_{2i}; \quad \mathbf{s}_{2(p \times p)} = \frac{1}{n_2 - 1} \sum_{i=1}^{n_2} (\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)(\mathbf{x}_{2i} - \bar{\mathbf{x}}_2)' \end{aligned} \quad (3.11)$$

Since it is assumed that the two groups came from one parent population and, hence, have the same covariance matrix  $\boldsymbol{\Sigma}$ , the sample covariance matrices are combined (pooled) to derive a single, unbiased estimator of  $\boldsymbol{\Sigma}$ . In particular the weighted average

$$\mathbf{S}_{pl} = \frac{(n_1 - 1)\mathbf{s}_1 + (n_2 - 1)\mathbf{s}_2}{(n_1 + n_2 - 2)} \quad (3.12)$$

is an unbiased estimate of  $\boldsymbol{\Sigma}$  if the data matrices  $\mathbf{X}_1$  and  $\mathbf{X}_2$  contain random samples from the populations  $\pi_1$  and  $\pi_2$ , respectively. Note that in order for  $\mathbf{S}_{pl}^{-1}$  to exist, we must have  $n_1 + n_2 - 2 > P$ .

Hence, the major procedures of discriminant analysis to be applied in this study are discriminant function analysis for two groups, test of significance, stepwise Variable selection, classification analysis and estimation of misclassification rates.

### 3.4.3.1 The Discriminant Function for Two Groups

Discriminant functions are linear combinations of variables that best separate groups. We assume that the two groups to be compared have the same covariance matrix  $\mathbf{\Sigma}$  but distinct mean vectors  $\boldsymbol{\mu}_1$  and  $\boldsymbol{\mu}_2$ . We work with samples  $\mathbf{X}_{1(p \times n_1)} = [\mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{1n_1}]$  and  $\mathbf{X}_{2(p \times n_2)} = [\mathbf{x}_{21}, \mathbf{x}_{22}, \dots, \mathbf{x}_{2n_2}]$ . The discriminant function is the linear combination of the  $p$  variables that maximize the distance between the two (transformed) group mean vectors. A linear combination  $y = \hat{\boldsymbol{\alpha}}' \mathbf{x}$  transforms each observation vector to scalar. Fisher (1936) used this linear combination to transform the multivariate observations  $\mathbf{x}$  to univariate observations  $y$  such that the  $y$ 's derived from populations  $\pi_1$  and  $\pi_2$  were separated as much as possible. We wish to find the vector  $\hat{\boldsymbol{\alpha}}$  that maximizes the standard difference

$$\frac{(\bar{y}_1 - \bar{y}_2)^2}{s_y^2} = \frac{[\hat{\boldsymbol{\alpha}}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)]^2}{\hat{\boldsymbol{\alpha}}' \mathbf{S}_{pl} \hat{\boldsymbol{\alpha}}} \quad (3.13)$$

$$\text{where } \bar{y}_1 = \hat{\boldsymbol{\alpha}}' \bar{\mathbf{x}}_1, \bar{y}_2 = \hat{\boldsymbol{\alpha}}' \bar{\mathbf{x}}_2 \text{ and } s_y^2 = \frac{\sum_{i=1}^{n_1} (y_{1i} - \bar{y}_1)^2 + \sum_{i=1}^{n_2} (y_{2i} - \bar{y}_2)^2}{n_1 + n_2 - 2} \quad (3.14)$$

The maximum of (3.13) occurs when

$$\hat{\boldsymbol{\alpha}} = \mathbf{S}^{-1}_{pl} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p)' \quad (3.15)$$

or when  $\hat{\boldsymbol{\alpha}}$  is any multiple of  $\mathbf{S}^{-1}_{pl} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ . When  $\hat{\boldsymbol{\alpha}} = \mathbf{S}^{-1}_{pl} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$  is used in  $y = \hat{\boldsymbol{\alpha}}' \mathbf{x}$ , the linear combination becomes

$$y = \hat{\boldsymbol{\alpha}}' \mathbf{x} = \mathbf{S}^{-1}_{pl} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{x} \quad (3.16)$$

and this function is called Fisher linear discriminant function. Since any multiple of  $\mathbf{S}^{-1}_{pl} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$  will maximize (3.13), the maximizing vector  $\hat{\boldsymbol{\alpha}}$  is not unique. However, its dimension is unique; that is the relative values or ratios of  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p$  are unique.

The optimum direction given by  $\hat{\mathbf{a}} = \mathbf{S}^{-1}_{pl}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$  is effectively parallel to the line joining the centroids of the two groups, that is,  $\bar{\mathbf{x}}_1$  and  $\bar{\mathbf{x}}_2$ . This can be seen by substituting (3.15) into (3.13) to obtain

$$\frac{(\bar{y}_1 - \bar{y}_2)^2}{s_y^2} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1}_{pl} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) = D^2 \quad (3.17)$$

where,  $D^2$  is the Mahalanobis squared distance between group mean vectors. Thus,  $(\bar{y}_1 - \bar{y}_2)^2 / s_y^2 = \hat{\mathbf{a}}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$ , and any other direction than that represented by  $\hat{\mathbf{a}} = \mathbf{S}^{-1}_{pl}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$  would yield a smaller difference between  $\hat{\mathbf{a}}\bar{\mathbf{x}}_1$  and  $\hat{\mathbf{a}}\bar{\mathbf{x}}_2$  (Alvin and Rencher, 1995).

From (3.13) the standard difference can also be written as

$$\frac{(\bar{y}_1 - \bar{y}_2)^2}{s_y^2} = \frac{[\hat{\mathbf{a}}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)]^2}{\hat{\mathbf{a}}' \mathbf{S}_{pl} \hat{\mathbf{a}}} = \frac{\hat{\mathbf{a}}'(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \hat{\mathbf{a}}}{\hat{\mathbf{a}}' \mathbf{S}_{pl} \hat{\mathbf{a}}} \quad (3.18)$$

To extend (3.18) to K groups, we use the  $\mathbf{H}$  matrix from MANOVA in place of  $(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'$  and  $\mathbf{E}$  in place of  $\mathbf{S}_{pl}$  to obtain the ratio

$$\Lambda = \frac{\hat{\mathbf{a}}' \mathbf{H} \hat{\mathbf{a}}}{\hat{\mathbf{a}}' \mathbf{E} \hat{\mathbf{a}}} \quad (3.19)$$

where  $\mathbf{H}$  is the between sum of squares and products matrix and  $\mathbf{E}$  is the within sum of squares and products matrix. After applying mathematical manipulation, the solutions of (3.19) are the eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_s$  and associated eigenvectors  $\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \dots, \hat{\mathbf{a}}_s$ , of  $\mathbf{E}^{-1}\mathbf{H}$ . We consider the eigenvalues to be ranked as  $\lambda_1 > \lambda_2 > \dots > \lambda_s$ . The number of (nonzero) eigenvalues  $s$  is the rank of  $\mathbf{H}$ , which can be found as the smaller of  $K-1$  or  $P$ . Thus the largest eigenvalue  $\lambda_1$  is the maximum value of (3.19). Hence, the discriminant function that maximally separates the means is  $y_1 = \hat{\mathbf{a}}_1' \mathbf{x}$ . From the  $s$  eigenvectors, we obtain  $s$  discriminant functions.

In our case, there are only two groups (K=2) and hence we have only one eigenvector  $\hat{\mathbf{u}}$  and its corresponding eigenvalue  $\lambda$ . Since  $s = \text{minimum}(K-1, P) = 1$ , we have only one discriminant function which is given as:

$$y = \hat{\mathbf{u}}' \mathbf{x} = \hat{\alpha}_0 + \hat{\alpha}_1 x_1 + \hat{\alpha}_2 x_2 + \dots + \hat{\alpha}_p x_p \quad (3.20)$$

### 3.4.3.2 Standardized Discriminant Function

In the two group case the relative contribution of X's to separation of the two groups can best be assessed by comparing the coefficients  $\alpha_j$ ,  $j = 1, 2, \dots, P$ , in equation (3.20).

The use of discriminant function to assess contribution of the X's to separation of groups gives meaningful interpretation only if the X's are commensurate, that is, measured on same scale and with comparable variances. If the X's are not commensurate, we need coefficients  $\alpha^*_j$  that are applicable to standardized variables. Hence, the standardized coefficients must be of the form

$$\alpha^*_j = s_j \alpha_j, j = 1, 2, \dots, P \quad (3.21)$$

where  $s_j$  is the within- sample standard deviation of the  $j^{\text{th}}$  variable obtained as the square root of the  $j^{\text{th}}$  diagonal element of  $\mathbf{S}_{pl}$ . In vector form, the standardized coefficient is given as:

$$\mathbf{\alpha}^* = (\text{diag } \mathbf{S}_{pl})^{1/2} \mathbf{\alpha} \quad (3.22)$$

Hence, the standardized discriminant function is obtained by substituting (3.22) into (3.20) and becomes

$$y = \hat{\mathbf{u}}^* \mathbf{x} = \hat{\alpha}_1^* Z_1 + \hat{\alpha}_2^* Z_2 + \dots + \hat{\alpha}_p^* Z_p \quad (3.23)$$

where  $Z_j$  ( $j = 1, 2, \dots, p$ ) is the standardized score of the  $j^{\text{th}}$  variable.

### 3.4.3.3 Tests of Significance in Discriminant analysis

In order to test hypotheses; we need the assumption of normality for groups. In this sub-section we discuss how to test four hypotheses that are mainly important in this study. These are test of equality of means between two groups, test of equality of covariance matrices, test of the significance of the discriminant function and measure of association of discriminant function.

### 3.4.3.3.1 Test of equality of means between two groups

We now consider the case where  $P$  variables are measured on each group. We wish to test

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{Versus} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

There are many test statistics to test this hypothesis. However, in this case we used the two most commonly used test statistics. These are Hotelling's  $T^2$ -test and Wilk's Lambda test.

Hotelling's  $T^2$ -test statistic, which is developed from the univariate t-test, is given by

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \quad (3.24)$$

which is distributed as  $T^2_{P, n_1+n_2-2}$  and  $H_0$  is rejected if  $T^2 > T^2_{\alpha, P, n_1+n_2-2}$ .

We reject  $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  if the standardized difference between  $\bar{\mathbf{x}}_1$  and  $\bar{\mathbf{x}}_2$  is large and hence  $T^2$  is large. This test can be readily transformed to an F-statistic as

$$\frac{n_1 + n_2 - P - 1}{(n_1 + n_2 - 2)P} T^2 = F_{P, n_1+n_2-P-1}$$

As a result, conclusion of the test is given based on the approximated F-value.

The second appropriate test is the based on Wilk's Lambda ( $\Lambda$ ) test statistic, which is given as:

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} \quad (3.25)$$

We reject  $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  for small value of  $\Lambda$ . The range of  $\Lambda$  is  $0 \leq \Lambda \leq 1$ , and the test based on Wilks'  $\Lambda$  is an inverse test in the sense that we reject  $H_0$  for small value of  $\Lambda$ . If the group mean vectors were equal, we would have  $\mathbf{H} = 0$ , and  $\Lambda = 1$ . On the other hand, as the sample mean vectors become more widely spread apart compared to the within-sample variation,  $\mathbf{H}$  becomes much larger than  $\mathbf{E}$ , and  $\Lambda$  approaches to zero.

Wilks'  $\Lambda$  can be expressed in terms of the eigenvalue  $\lambda$  of  $\mathbf{E}^{-1}\mathbf{H}$ , and is given as

$$\Lambda = \frac{1}{1 + \lambda} \quad (3.26)$$

Similarly, the Wilks'  $\Lambda$  can also be approximated by F-distribution. But the derivation is complicated and only done by statistical softwares. Conclusion is made based on the F-value.

Finally, if the hypothesis  $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  is rejected, the implication is that  $\mu_{1j} = \mu_{2j}$  will be rejected for at least one  $j = 1, 2, \dots, p$ . But, there is no guarantee that  $\mu_{1j} = \mu_{2j}$  will be rejected for all  $j$  (Rencher, 1995). Further test of equality of group means for each variable is made by a univariate-ANOVA. This method follows the same procedure in using Wilks'  $\Lambda$  discussed above but the test is made separately for each variable.

There are many procedures that could be used to check the contribution of each variable following rejection of  $H_0$  by using either of the above tests. In this study, however, we use the standardized discriminant function coefficients to see the effect of each variable after  $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  is rejected. The standardized discriminant function coefficients procedure compares the (absolute value of) coefficients in the discriminant function to find the effect of each variable in separating the two groups.

### 3.4.3.3.2 Test of Equality of Covariance Matrices of Two Groups

In this case we need to discuss how to test whether the covariance matrices of the two groups are equal or not. The hypothesis to be tested is

$$H_0 : \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 \quad \text{Versus} \quad H_1 : \boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$$

We assume the two groups of size  $n_1$  and  $n_2$  were taken from multivariate normal distributions.

For this test, we use Box's M test which is given by

$$M = \frac{|\mathbf{s}_1|^{v_1/2} |\mathbf{s}_2|^{v_2/2}}{|\mathbf{S}_{pl}|^{\sum v_i/2}} \quad (3.28)$$

in which  $v_i = n_i - 1$ ,  $\mathbf{s}_i$  is the covariance matrix of the  $i^{th}$  group ( $i = 1, 2$ ), and  $\mathbf{S}_{pl}$  is the pooled sample covariance matrix.

Box (1949, 1950: cited in Rencher, 1995) gave  $\chi^2$  and  $F$  approximations for the distribution of  $M$ . However, derivations of the approximations are complicated. Some statistical packages such as SPSS, STATA or SAS give the approximation results. Hence, we use the F-approximation results for conclusion.

### 3.4.3.3.3 Measure of Association for Discriminant Function

Measure of association between the response variable and the  $p$  independent variables is given by

Roy's statistic ( $\eta^2$ ) as  $\eta^2 = \frac{\lambda}{1 + \lambda}$ , where  $\lambda$  is the eigenvalue. (3.29)

This measure attempts to answer the question 'How well do the variables separate the groups?' It

is noted that Roy's statistic serves as  $R^2$ -like measure of association in regression analysis. It can

be shown that the square root of the quantity in (3.29),  $\eta = \sqrt{\frac{\lambda}{1 + \lambda}}$  is the maximum correlation

between a linear combination of the  $p$  independent variables and the dependent variable. This type of correlation is often called a canonical correlation (Rencher, 1995). There is no rule of thumb for the canonical correlation. However, it can be tested for significance using the same procedure for testing the significance of eigenvalue given below.

### 3.4.3.3.4 Test of Significance of the Discriminant function

In this case we need to test the significance of the eigenvalue and thereby the discriminant functions. This test is similar with the test of equality of group means. In fact, rejecting the null

hypothesis of equal means implies the discriminant function,  $y = \hat{\boldsymbol{\alpha}}' \mathbf{x}$ , is significant. That is, if

$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$  is rejected, we conclude that the eigenvalue ( $\lambda$ ) is significantly different from zero.

Since  $\lambda$  is the only largest eigenvalue, one can be sure of its significance, along with that of  $y = \hat{\boldsymbol{\alpha}}' \mathbf{x}$ .

### 3.4.3.4 Stepwise Selection of Variables

In many applications, there are large numbers of independent variables that affect the response variable. This is because the researcher is unsure as to which variable is important. This study is

limited to procedures that remove or add variables one at a time so that a subset of the  $p$  variables, which best discriminate among the two mutually distinct groups, will be selected using stepwise methods.

If there are no variables that we have a priori interest in testing for significance, we can do a data-directed search for the variables that best separate the households into food secure and food insecure. Such a strategy is often called *stepwise discriminant analysis*. The procedure appears in many software packages. For example in SPSS, there are different stepwise selection methods. Among these selection methods the Wilks' Lambda stepwise variable selection approach is used in this study.

We first describe an approach that is usually called *forward selection*. At the first step we calculate  $\Lambda ( X_j )$  for each individual variable and choose the one with minimum  $\Lambda ( X_j )$  (or maximum associated  $F$ ). At the second step, we calculate  $\Lambda ( X_j | X_1 )$  for each of the  $P-1$  variables not entered at the first step, where  $X_1$  indicate the first variable entered. For the second variable we choose the one with minimum  $\Lambda ( X_j | X_1 )$  (or maximum associated partial  $F$ ), that is, the variable that adds the maximum separation to the one entered in step 1. Denote the variable entered in step 2 by  $X_2$ . Continue this process until the  $F$  falls below some predetermined threshold value, say  $F_{in}$ . Similarly, *backward elimination* is a variable selection technique in which the variable that contributes least is deleted at each step, as indicated by the partial  $F$ .

*Stepwise selection* is a combination of the forward and backward approaches. Variables are selected one at a time, and at each step. The variables already selected would be reexamined to see if each still constitutes a significant amount. In this study we used the variables which are final output of the stepwise variable selection technique as major factors causing household food security in the study area.

### 3.4.3.5 Classification

In classification, a sampling unit (subject or object) whose group membership is unknown is assigned to a group on the basis of the vector of  $P$  measured values,  $\mathbf{x}$ , associated with the unit. There are many principles of allocation of observation to a certain group (class). However in our case, we use the Fisher (1936) based allocation procedure and the optimal ECM procedure.

To use the Fisher classification procedure, the principal assumption is that the two populations have the same covariance matrix. Normality is not required. We obtain a sample from each of the two populations and compute  $\bar{\mathbf{x}}_1$ ,  $\bar{\mathbf{x}}_2$  and  $\mathbf{S}_{pooled}$ . A simple procedure for classification can be based on the discriminant function given in (3.16), that is,  $y = \hat{\mathbf{a}}' \mathbf{x} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} \mathbf{x}$ . The midpoint,  $\hat{m}$ , between the two univariate sample means,  $\bar{y}_1 = \hat{\mathbf{a}}' \bar{\mathbf{x}}_1$  and  $\bar{y}_2 = \hat{\mathbf{a}}' \bar{\mathbf{x}}_2$  is given by

$$\hat{m} = \frac{1}{2}(\bar{y}_1 + \bar{y}_2) = \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \quad (3.30)$$

Hence, an allocation rule based on Fisher's sample discriminant function is given as follows:

Allocate  $\mathbf{x}_0$  to  $\pi_1$  if

$$y_0 \geq \hat{m}, \text{ where } y_0 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} \mathbf{x}_0$$

Otherwise allocate  $\mathbf{x}_0$  to  $\pi_2$ .

This classification rule employs the same discriminant function used in (3.20). Thus in the two-group case the discriminant function serves as a linear classification function. However, in the several-group case, the classification functions are different from the descriptive discriminant functions (Rencher, 1995).

The other method of classification is based on the optimal ECM. The optimal allocation, the allocation that minimizes ECM, rule can be modified by substituting the sample estimates for the population parameters in (3.9) and gives the sample classification rule as:

Allocate  $\mathbf{x}_0$  to  $\pi_1$  if

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} \mathbf{x}_0 - \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_{pooled}^{-1} (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \geq \ln \left[ \frac{C(1|2)}{C(2|1)} \right] \left[ \frac{P_2}{P_1} \right]$$

Otherwise allocate  $\mathbf{x}_0$  to  $\pi_2$ .

If  $[C(1|2)/C(2|1)](P_2/P_1)=1$  so that  $\ln [C(1|2)/C(2|1)](P_2/P_1)=0$  then this rule is comparable to the rule based on Fisher's linear discriminant function. That is, Fisher's classification rule is equivalent to the minimum ECM rule with equal prior probabilities and equal cost of misclassification provided that the two normal populations have the same covariance matrix.

### 3.4.3.6 Estimating Misclassification rates

To judge the ability of classification procedure to predict group membership we usually use the probability of misclassification rate. A simple estimate of the error rate can be obtained by trying out the classification procedure on the same data set that has been used to compute the classification function. This method is commonly referred to as resubstitution. Each observation vector is submitted to the classification function and assigned to a group. We then count the number of correct classifications and misclassifications. The proportion of misclassifications resulting from resubstitution is called apparent error rate (AER).

Let among the  $n_1$  observations in  $\pi_1$ ,  $n_{11}$  are correctly classified in to  $\pi_1$  and  $n_{12}$  are misclassified into  $\pi_2$ , where  $n_1=n_{11}+n_{12}$ . Similarly, of the  $n_2$  observations in  $\pi_2$ ,  $n_{21}$  are misclassified in to  $\pi_1$  and  $n_{22}$  are classified in to  $\pi_2$ , where  $n_2=n_{21}+n_{22}$ . Thus, Apparent Error Rate (AER) is given as

$$\text{AER} = \frac{n_{12} + n_{21}}{n_1 + n_2} \quad (3.31)$$

Similarly, we can define as, Apparent correct classification rate = 
$$\frac{n_{11} + n_{22}}{n_1 + n_2} \quad (3.32)$$

Clearly,  $\text{AER} = 1 - \text{apparent correct classification rate}$ .

Apparent error rate is an estimate of the probability that our classification function based on the present sample will misclassify a future observation. This probability is called the actual error rate. An error rate of less than 0.30 is a more realistic estimate of what the classification functions can do with future samples (Rencher, 1995).

# CHAPTER FOUR

## SATATISTICAL DATA ANALYSIS

The data analysis is done using SPSS, STATA and SAS statistical (software) packages. The results of the analysis are divided into six sections: Evaluation of assumptions, descriptive analysis results, bivariate analysis results, stepwise variable selection, results of discriminant analysis and results of classification. These results and their discussions are presented below.

### 4.1 Evaluation of Assumptions

The major assumptions of discriminant analysis are the multivariate normality and the homogeneity of within covariance matrices between the two groups. The normality assumption helps to decide whether parametric or non-parametric discriminant analysis is appropriate to analyse the data at hand. On the other hand, the homogeneity assumption of covariance matrices helps to choose whether linear discriminant analysis or quadratic discriminant analysis is appropriate for the data.

#### 4.1.1 Checking Multivariate Normality

The assumption of multivariate normality states that the predictor variables should fulfill the multivariate normality in each group. The commonly used techniques for detecting multivariate normality are the quantile-quantile (Q-Q) and the Chi-square plots. The Q-Q plot displays the ordered data of the discriminant score quantiles against normal quantiles while the Chi-square plot displays the Mahalanobis distance versus the Chi-square quantiles with thirteen degrees of freedom. Multivariate normality is said to be met if both plots provide scatter plots which closely lay on the line that pass through the origin.

The Q-Q plot and the Chi-square plots presented in ANNEX C show that the data on the rural households of the Tigray region approximately satisfied the assumption of multivariate normality. Similarly, the chi-square plots indicate that both groups have approximate normal distributions. When there are greater than 20 observations in the smallest group, the test is robust to violations of multivariate normality, usually if violation is caused by skewness and not outliers (Fidell and

Tabachnick, 2007). In our case the sample sizes of the two groups are  $n_1=350$  (number of food insecure households) and  $n_2=478$  (number of food secure households), which make the discriminant analysis sufficiently robust that we do not worry about moderate departures from normality.

The test of multivariate normality can also be supported by checking the degree of skewness and kurtosis for each continuous variable. The calculated values of these quantities are given in Table B2, Annex B. These coefficients clearly show that each variable has a normal distribution except CDR and EDUC. However, individual normality doesn't necessarily imply multivariate normality (Johnson, 2002). Hence, we should not worry when some predictors violate the normality assumption if the multivariate normality is met.

#### 4.1.2 Testing Homogeneity of Covariance Matrices between Groups

This assumption states that the variance-covariance matrix of the predictor variables is the same in both groups so they can be pooled together to estimate the error variance.

The null hypothesis to be tested is that the covariance matrices of the two groups are the same in the population. **Box's M** test is used to test this hypothesis and the SPSS output is given in Table 4.1 below

Table 4.1: Box's Test of Equality of Covariance Matrices

|         |         |             |
|---------|---------|-------------|
| Box's M |         | 44.538      |
| F       | Approx. | 1.224       |
|         | df1     | 36          |
|         | df2     | 1906022.843 |
|         | Sig.    | .167        |

Tests null hypothesis of equal population covariance matrices.

The Box's M result shows that there is no significant difference between the covariance matrices of the two food security status groups (approximate  $F=1.224$ ,  $p > 0.05$ ) and we do not reject the null hypothesis. Thus, the assumption of homogeneity of covariance matrices between the two groups is met.

Based on the above two procedures, it can be concluded that the assumptions of multivariate normality and homogeneity of covariance matrices were sufficiently satisfied. Hence, the appropriate model for analyzing the data on the rural households included in this study and for discriminating these two household groups based on the thirteen attributes is the parametric linear discriminant analysis.

#### 4.1.3 Checking outliers

An outlier is a case with such an extreme value on one variable (a univariate outlier) or such a strange combination of scores on two or more variables (multivariate outlier) that distorts statistics. Outliers can be found in both dichotomous and continuous variables and they lead to both Type I and Type II errors (Fidell and Tabachnick, 2007). Hence, it is better to check the presence of outliers before proceeding to the analysis.

The univariate outliers were checked for each variable using Box plot. It was found that there was no extreme value in each variable and, hence, there is no problem of univariate outliers. Moreover, the 828 cases were also screened for multivariate outliers using SPSS. Cases with standard residual greater than 3 or less than -3 are expected to cause multivariate outliers and, hence, they are selected and their associated Mahalanobis distances are calculated.

The criterion for multivariate outliers is Mahalanobis distance at  $p < 0.001$  (Fidell and Tabachnick, 2007). Mahalanobis distance is evaluated as  $\chi^2$  with degrees of freedom equal to the number of variables, (in our case thirteen). Any case with a Mahalanobis distance greater than  $\chi^2(13) = 34.528$  is a multivariate outlier. In our study, all cases have Mahalanobis distance less than 34.528. Hence, there is no problem of multivariate outliers in our data.

#### 4.1.4 Checking Multicollinearity

Multicollinearity or singularity may occur with highly redundant predictors, making matrix inversion impossible. Analysis of data with the presence of collinearity among predictors may give unreliable results. Hence, we need to check the presence of multicollinearity before proceeding to the analysis. Some of the appropriate methods for assessing multicollinearity are variance inflation

factor (VIF) and Condition index. The STATA and SAS outputs for multicollinearity information are given in ANNEX B.

Variance Inflation Factors (VIF) greater than 10 are generally seen as indicative of severe multicollinearity. The 1/VIF column is the tolerance and it ranges from 0 to 1, with 1 being the absence of multicollinearity. In our case all of the VIFs are below 4 and all of the tolerances are close to one indicating that there is no problem of multicollinearity in our data. Similarly, from the eigenvalues and condition indexes no severe problems of multicollinearity were noted except for the last two variables. As a rule of thumb a condition index below 15 is indicative of the absence of multicollinearity problem (Gujarati, 1988).

## **4.2 Descriptive Analysis Results**

Based on the calorie requirement of 2100 Kcal per day per person, out of the 828 sample rural households of the Tigray regional state 350 (42.27%) and 478 (57.73%) were found food insecure and food secure households, respectively. The head count ratio, average food insecurity gap and squared food insecurity gap are called food insecurity indices and are presented in Table 4.2.

The head-count ratio measures the incidence of food insecurity, the proportion of households defined to be food insecure. In this study head count index was found to be about 42.3% indicating that, on average, 42.3% of the rural households in Tigray region are unable to meet the stipulated minimum level of calorie intake i.e. 2100 Kcal per adult equivalence per day.

The food insecurity gap measures the mean depth of food insecurity among the food insecure households. It is the mean proportion by which the food security status of the food insecure households falls below the minimum level of calorie requirement. In our case the food insecurity gap indicates that the food insecure households are 32.5% far off from the minimum level of calorie requirement i.e. 2100 Kcal.

And the squared food insecurity gap measures the severity of food insecurity among the food insecure households. It gives more weight to the average income shortfall of the most food insecure of the food insecure households. Thus, it measures the squared proportional shortfalls

from the minimum level of calorie intake. The drawback of this index of food insecurity is that it is not easy to interpret (Sen, 1981). However, we can say that the severity of food insecurity in the rural areas of Tigray region is about 18%.

Table 4.2: Food Insecurity Indices

| Measures of Food insecurity | value  |
|-----------------------------|--------|
| Head count ratio            | 42.27% |
| Average Food insecurity gap | 32.50% |
| Squared Food insecurity gap | 18.08% |

The group statistics given in Table 4.3 provide summaries of descriptive statistics such as means and standard deviations of variables which are expected to determine the household food security status. From the group statistics it can be observed that food insecure households have greater averages of household size, child dependency ratio, distance to input sources and average monthly expenditure of the household than the food secured households.

Table 4.3: Group Statistics for continuous variables

| Variables     | Food Insecure HHs |          | Food secure HHs |          | Total   |          |
|---------------|-------------------|----------|-----------------|----------|---------|----------|
|               | $n_1=350$         |          | $n_2=478$       |          | $n=828$ |          |
|               | Mean              | Std.dv   | Mean            | Std.dv   | Mean    | Std.dv   |
| <b>HHSZ</b>   | 5.5314            | 2.14593  | 4.4205          | 2.20941  | 4.89    | 2.24955  |
| <b>AGE</b>    | 47.6057           | 15.74689 | 48.2050         | 15.73429 | 47.95   | 15.73288 |
| <b>EDUC</b>   | 1.1086            | 2.30974  | 1.657           | 2.28877  | 1.44    | 2.31331  |
| <b>DIST</b>   | 13.6949           | 5.64939  | 8.0479          | 5.45927  | 10.44   | 6.20079  |
| <b>TLU</b>    | 3.2168            | 2.27749  | 4.4762          | 2.27051  | 3.94    | 2.35582  |
| <b>CDR</b>    | 1.3629            | 0.99613  | 1.0726          | 0.95234  | 1.20    | 0.98104  |
| <b>OXEN</b>   | 0.7714            | 0.90163  | 1.2803          | 0.99520  | 1.07    | 0.98874  |
| <b>FLSZ</b>   | 0.395             | 0.21862  | 0.5317          | 0.21959  | 0.47    | 0.22923  |
| <b>EXPEND</b> | 178.207           | 6.13606  | 172.847         | 6.3069   | 175.11  | 6.23714  |

HHs: Households; St.dv: Standard deviation

On the other hand, the food secure households have relatively greater averages on age of head of household, number of livestock, number of oxen and farm land size than their counter part households. We can also see that the group standard deviations are approximately equal, indicating that there is homogeneity of variances between the two household groups for each predictor.

### **4.3 Bivariate Analysis Results**

The association between each explanatory variables and household food security status is conducted by cross-tabulating each predictor variables against the outcome variable. Moreover, a univariate analysis of variance (ANOVA) of each predictor variables against the household food security status is performed to identify the significant candidate predictor variables that would qualify for the multivariate discriminant analysis.

The major factors that are expected to determine household food security status were first analyzed by considering the relationship of each predictor variable with the outcome variable. Table 4.4 provides the association of each predictor variable and household food security status. Based on the results given in this table, ten of the thirteen explanatory variables considered in this study were found statistically significantly associated with the status of household food security ( $p < 0.05$ ). They are household size, gender of head of the household, age of head of the household, educational level of head of the household, distance to the nearest input sources, use of fertilizer by the household, number of livestock, child dependency ratio, oxen ownership and farm land size.

When we see the summary bivariate results of the demographic predictors presented in Table 4.4, it was found that 502 of the sample households have 5 or less than 5 members out of which 34.7 % were found food insecure. On the other hand, out of the 326 households who have greater than 5 members 54 % were found to be food insecure. As a result, household size has a statistically significant association with household food security status ( $p < 0.05$ ). Based on the sampled rural households in Tigray region, out of the 828 households 70% are male headed households and the remaining 30% are female headed households. The proportion of food insecure households among the male headed households is about 39% which is lower as compared with the proportion of food insecure households among the female headed households which is 49%. This supports the idea that gender of head of a household has statistically significant association with food security status. Moreover, age of head of household has significant relationship with food security status in that

the proportion of food insecure households is higher in the age group between 35 and 46 years than the other age groups. The last demographic predictor that has strongly related with food security status is child dependency ratio (CDR). The proportion of food insecurity is higher among households with CDR greater than or equal to the sample mean (1.20).

When we come to the socio-economic predictors, we find that about 42 % of the sampled household heads can read and/or write and the remaining 58% can not read and write. Among the household heads that can read/write 28.9 % are food insecure while about 52% of the household heads who can not read and write are food insecure. These figures show that educational level of head of a household has statistically significant impact on the household food security status. From the results in Table 4.4, it can also be observed that 51% of the households travel less than 10 Km to the input sources and facility centers and the remaining households travel greater than 10 Km. Out of the households who travel less than 10 Km, 25.1 % households are food insecure while about 60% of the households who travel greater than 10 km are food insecure. That is the proportion of food insecure is higher among households who travel above 10km than those who travel less than 10km. Hence, distance to be traveled by the household to input source is strongly related with household food security status.

About 70.8 % of the sampled households are fertilizer users out of which about 39% are food insecure, and from those who do not use fertilizer, 50.4 % were found food insecure. The association between use of fertilizer and food security status is statistically significant ( $p < 0.05$ ). As depicted in table 4.4, livestock ownership is also one of the major factors that have significant impact on food security status. The proportion of food insecure households is higher (50.4%) with households who own less than the sample mean of livestock (3.94 TLU) than the households who own greater than the sample mean of livestock.

Lastly we see that about 35% of the sampled households do not own oxen for plowing their farm land. The proportion of food insecure households among the households with no ox is 61.2%, which is very high as compared to the proportion of food insecure households among the ox owner households (32%). It can also be seen that the proportion of food insecure is higher among households with smaller farm land size than the average farmland size of the sample (0.47 hectare)

Table 4.4: Association between HFS and the predictor variables

| Variables        | Number of HHs | Total % | Food insecure % | Food secure % | Pearson Chi-square (p-value) | df |
|------------------|---------------|---------|-----------------|---------------|------------------------------|----|
| <b>HHSZ</b>      |               |         |                 |               |                              |    |
| ≤ 5              | 502           | 60.6    | 34.7            | 65.3          | 30.25<br>(0.000)*            | 1  |
| ≥ 6              | 326           | 39.4    | 54              | 46            |                              |    |
| <b>GEND</b>      |               |         |                 |               |                              |    |
| Male             | 578           | 69.8    | 39.3            | 60.7          | 7.047<br>(0.008)*            | 1  |
| Female           | 250           | 30.2    | 49.2            | 50.8          |                              |    |
| <b>AGE</b>       |               |         |                 |               |                              |    |
| < 35 years       | 191           | 23.1    | 40.8            | 59.2          | 8.078<br>(0.044)*            | 3  |
| 35-46            | 221           | 26.7    | 50.2            | 49.8          |                              |    |
| 47-59            | 200           | 24.2    | 39              | 61            |                              |    |
| ≥ 60             | 216           | 26.1    | 38.4            | 61.6          |                              |    |
| <b>EDUC</b>      |               |         |                 |               |                              |    |
| Can read/write   | 346           | 41.8    | 28.9            | 71.1          | 43.53<br>(0.000)*            | 1  |
| Can't read/write | 482           | 58.2    | 51.9            | 48.1          |                              |    |
| <b>DIST</b>      |               |         |                 |               |                              |    |
| ≤ 10 Km          | 422           | 51      | 25.1            | 74.9          | 103.75<br>(0.000)*           | 1  |
| > 10 Km          | 406           | 49      | 60.1            | 39.9          |                              |    |
| <b>AGREXT</b>    |               |         |                 |               |                              |    |
| Use              | 537           | 64.9    | 43.4            | 56.6          | 0.784<br>(0.376)             | 1  |
| Do not use       | 291           | 35.1    | 40.2            | 59.8          |                              |    |
| <b>FRTLZR</b>    |               |         |                 |               |                              |    |
| Use              | 586           | 70.8    | 38.9            | 61.1          | 9.291<br>(0.002)*            | 1  |
| Do not use       | 242           | 29.2    | 50.4            | 49.6          |                              |    |
| <b>TLU</b>       |               |         |                 |               |                              |    |
| ≥ 3.94           | 403           | 48.7    | 33.7            | 66.3          | 23.37<br>(0.000)*            | 1  |
| < 3.94           | 425           | 51.3    | 50.4            | 49.6          |                              |    |
| <b>CDR</b>       |               |         |                 |               |                              |    |
| < 1.20           | 478           | 57.7    | 36.4            | 63.6          | 15.96<br>(0.000)*            | 1  |
| ≥ 1.20           | 350           | 42.3    | 50.3            | 49.7          |                              |    |
| <b>OXEN</b>      |               |         |                 |               |                              |    |
| Own Ox           | 537           | 64.9    | 32              | 68            | 65.66<br>(0.000)*            | 1  |
| No Ox            | 291           | 35.1    | 61.2            | 38.8          |                              |    |
| <b>FLSZ</b>      |               |         |                 |               |                              |    |
| ≥ 0.47 Hectare   | 380           | 45.9    | 29.5            | 70.5          | 47.13<br>(0.000)*            | 1  |
| < 0.47 Hectare   | 448           | 54.1    | 53.1            | 46.9          |                              |    |
| <b>EXPEND</b>    |               |         |                 |               |                              |    |
| ≥ 175.0 Birr     | 370           | 44.7    | 41.4            | 58.6          | 0.232<br>(0.63)              | 1  |
| < 175.0 Birr     | 458           | 55.3    | 43              | 57            |                              |    |
| <b>SEED</b>      |               |         |                 |               |                              |    |
| Use              | 360           | 43.5    | 41.9            | 58.1          | 0.028<br>(0.868)             | 1  |
| Do not use       | 468           | 56.5    | 42.5            | 57.5          |                              |    |

\* Significant (p<0.05)

We note that all predictor variables included in this study are categorized merely for the purpose of checking the association between the household food security status and each predictor variable. In other analyses of this study, the continuous and discrete variables are taken as they are originally.

#### 4.4 Results of stepwise variable selection

Up to now all of the thirteen predictors under study were assumed as discriminating variables for discrimination and classification purposes. Probably the most common application of discriminant analysis is to select a subset of the original predictor variables that provide maximum discrimination between the two household groups based on their status of food security.

This study used Wilks' lambda stepwise procedure discussed in section 3.4 to select the best subset of variables for classification of households with minimum tolerance of F-to-enter = 0.05 and F-to-remove=0.10. In each step the stepwise discriminant analysis output produces the Wilks' lambda and the associated F-to-enter and F-to-remove for variables not in the analysis and variables in the analysis. At each step, the variable that minimizes the overall Wilks' lambda or the variable with the largest F-to-enter is included in the analysis. The selected variables are given in Table 4.5.

Table 4.5 Variables selected by stepwise variable selection procedure

| Step |        | Wilks' Lambda |     |     |     |           |     |     |      |
|------|--------|---------------|-----|-----|-----|-----------|-----|-----|------|
|      |        | Statistic     | df1 | df2 | df3 | Exact F   |     |     |      |
|      |        |               |     |     |     | Statistic | df1 | df2 | Sig. |
| 1    | DIST   | .797          | 1   | 1   | 826 | 209.90    | 1   | 826 | .000 |
| 2    | FLSZ   | .735          | 2   | 1   | 826 | 148.80    | 2   | 825 | .000 |
| 3    | OXEN   | .698          | 3   | 1   | 826 | 119.03    | 3   | 824 | .000 |
| 4    | HHSZ   | .646          | 4   | 1   | 826 | 112.86    | 4   | 823 | .000 |
| 5    | TLU    | .623          | 5   | 1   | 826 | 99.68     | 5   | 822 | .000 |
| 6    | FRTLZR | .615          | 6   | 1   | 826 | 85.49     | 6   | 821 | .000 |
| 7    | GEND   | .610          | 7   | 1   | 826 | 74.87     | 7   | 820 | .000 |
| 8    | EDUC   | .606          | 8   | 1   | 826 | 66.58     | 8   | 819 | .000 |

Based on the stepwise discriminant analysis results presented in Table 4.5 eight of the thirteen variables are selected for analysis to be included in the model and the remaining are excluded from analysis and, hence, omitted from the model. The variables which passed the stepwise variable selection procedure as candidate to be included in the model are: distance in km to input sources, farm land size, number of oxen, household size, livestock ownership, use of fertilizer, gender of head of household head and level of education of household head

## 4.5. Results of Discriminant Analysis and Discussion

Following the fulfillment of the assumptions, LDA was found the most appropriate model to classify the rural households of Tigray region into food secure and food insecure groups based on the discriminating variables. Hence, we use SPSS to compute the linear discriminant function coefficients which help for discriminating and classifying the households to the two food security status. But first it is better to begin with the test of significance of group means.

### 4.5.1 Test of Equality of group means

Over all equality of group means is, first, tested by applying MANOVA and its results are presented in Table 4.6. As discussed in the methodology part, the values of Hotelling's and the Wilk's lambda statistics can be approximated by F-value. The results depicted in Table 4.6 show that the group means are significantly different ( $F= 42.20$ ,  $p<0.0001$ ). This implies that the two household groups have significantly different means.

Table 4.6: MANOVA for Testing Equality of Group means

| Source   | Statistic          | Value  | df  | F(df1, df2) = | F   | Sig. |        |
|----------|--------------------|--------|-----|---------------|-----|------|--------|
| HFS      | Wilk's lambda      | 0.6031 | 1   | 13            | 814 | 41.2 | 0.000* |
|          | Hotelling's- $T^2$ | 0.3969 |     | 13            | 814 | 41.2 | 0.000* |
|          | Pillai's trace     | 0.6580 |     | 13            | 814 | 41.2 | 0.000* |
|          | Roy's root         | 0.6580 |     | 13            | 814 | 41.2 | 0.000* |
| Residual |                    |        | 826 |               |     |      |        |
| Total    |                    |        | 827 |               |     |      |        |

\* Significant ( $p<0.05$ )

Furthermore, we use one-way ANOVA to check whether or not the two household groups are significantly different with respect to the mean of a particular variable. The one-way ANOVA statistical test of equality of group means is presented in Table 4.7. In the ANOVA table, values of Wilks' lambda and calculated F-statistics are given which are important to show significance of difference of the group means. The smaller the Wilks' lambda, the more important the independent variable to the discriminant function, and the F test of Wilks' lambda shows which variables have significant contribution to the discrimination of the two groups. The Wilks' lambda is relatively smaller for the variables such as distance to input sources, farm land size, livestock ownership, number of oxen and household size in comparison to the other variables which are almost one.

Table 4.7: Univariate ANOVA for the Test of Equality of Group Means

| <b>Variables</b> | <b>Wilks' Lambda</b> | <b>F</b> | <b>df1</b> | <b>df2</b> | <b>Sig.</b> |
|------------------|----------------------|----------|------------|------------|-------------|
| <b>HHSZ</b>      | 0.940                | 52.336   | 1          | 826        | .000*       |
| <b>GEND</b>      | 0.991                | 7.090    | 1          | 826        | .008*       |
| <b>AGE</b>       | 1.000                | 0.293    | 1          | 826        | .588        |
| <b>EDUC</b>      | 0.985                | 12.311   | 1          | 826        | .000*       |
| <b>DIST</b>      | 0.797                | 209.900  | 1          | 826        | .000*       |
| <b>AGREXT</b>    | 0.999                | 0.782    | 1          | 826        | .377        |
| <b>FRTLZR</b>    | 0.989                | 9.373    | 1          | 826        | .002*       |
| <b>OXEN</b>      | 0.930                | 62.002   | 1          | 826        | .000*       |
| <b>CDR</b>       | 0.979                | 18.050   | 1          | 826        | .000*       |
| <b>TLU</b>       | 0.935                | 57.163   | 1          | 826        | .000*       |
| <b>FLSZ</b>      | 0.913                | 78.596   | 1          | 826        | .000*       |
| <b>SEED</b>      | 1.000                | 0.028    | 1          | 826        | .868        |
| <b>EXPEND</b>    | 0.998                | 1.493    | 1          | 826        | .222        |

\* significant ( $p < 0.05$ ); df=degree of freedom

By the F test in Table 4.7, all variables which were selected by the stepwise variable selection have significant difference between group means. This should not be surprising since the stepwise discriminant variable selection guarantees to provide the significant variables.

## 4.5.2 Significance of the Discriminant Function

To test the significance of the discriminant function, one can test the number of eigenvalues that add significantly to the discrimination between groups. Since the dependent variable in this case has only two categories, there is only one discriminant function, which in turn has one eigenvalue. Testing the significance of the eigenvalue to the discrimination of the groups is the same as testing significance of the discriminant function.

We noted that when there are two groups, one discriminant function can be extracted from the data and its associated eigenvalue is given as  $\lambda = BSS/WSS$ , where BSS = between groups sum of squares and WSS= within group sum of squares. BSS=0 implies  $\lambda=0$  and that model has no discriminatory power. The larger the value of  $\lambda$ , the greater the discriminatory power of the model.

Based on the results given in Table 4.8, the eigenvalue of the discriminant function in our model is 0.650. It can also be observed that 100% of the variance is explained by the discriminant function. As a result, the cumulative percentage of the variance explained by the discriminant function is 100%. The canonical correlation ( $\eta = \sqrt{\lambda/(1 + \lambda)}$ ) between the predictor variables and the discriminant scores produced by the discriminant function is 0.628.

Table 4.8: Eigenvalues

| Function | Eigenvalue        | % of Variance | Cumulative % | Canonical Correlation |
|----------|-------------------|---------------|--------------|-----------------------|
| 1        | .650 <sup>a</sup> | 100.0         | 100.0        | .628                  |

a. First 1 canonical discriminant functions were used in the analysis.

To test the hypothesis  $H_0: \lambda = 0, \eta = 0$ , we use the Wilks' lambda statistic given in Table 4.9.

Based on the output, the chi-square test of the Wilks' lambda is significant ( $\chi^2=411.816$ ,

$p < 0.0001$ ). Hence, the null hypothesis is rejected and we conclude that the eigenvalue of the discriminant function used in this study is significantly different from zero. At the same time the canonical correlation was found statistically significant and hence there is good correlation between the explanatory variables and the discriminant scores. Significance of the eigenvalue and the canonical correlation imply that the discriminant function is statistically significant. Hence, the discriminant function best discriminates the two household groups based on the predictor variables included in this study.

Table 4.9: Wilks' lambda

| Test of Function(s) | Wilks' Lambda | Chi-square | df | Sig. |
|---------------------|---------------|------------|----|------|
| 1                   | .606          | 411.816    | 8  | .000 |

Since the discriminant function was found significant, its associated statistics are used in the interpretation of the effect of predictor variables on household food security status.

### 4.5.3 Discriminant Function and its Coefficients

In the methodology part it was discussed that there is only one discriminant function for two group discriminant analysis and, hence, we have one discriminant function in this study.

The discriminant function is created as a linear combination of the independent variables and is given as:  $y = \hat{\alpha}_0 + \hat{\alpha}_1x_1 + \hat{\alpha}_2x_2 + \dots + \hat{\alpha}_8x_8$  where the  $\hat{\alpha}$ 's are discriminant coefficients,  $x_j$  ( $j = 1, 2, \dots, 8$ ) is the selected independent variable and  $\alpha_0$  is a constant. This is analogous to the multiple regression, but the  $\hat{\alpha}$ 's are discriminant coefficients which maximize the distance between the means of the dependent variable.

Households get separate discriminant function scores for the discriminant function when their own scores on predictors are inserted into the equation.

The discriminant function score for the  $i^{th}$  ( $i = 1, 2, \dots, 828$ ) household using the unstandardized coefficients and the raw score of the predictors is given as

$$D_i = \hat{\alpha}_0 + \hat{\alpha}_1x_{i1} + \hat{\alpha}_2x_{i2} + \dots + \hat{\alpha}_p x_{ip}$$

where  $x_{ij}$  ( $i = 1, 2, \dots, 828, j = 1, 2, \dots, 8$ ) is raw score of the  $j^{th}$  variable on the  $i^{th}$  household.

The discriminant score ( $D_i$ ) can also be obtained by taking the standardized coefficients and the standardized score of the predictors as

$$D_i = \hat{\alpha}_1^* Z_{1i} + \hat{\alpha}_2^* Z_{2i} + \dots + \hat{\alpha}_p^* Z_{pi};$$
 where  $\alpha_j^*$  ( $j = 1, 2, \dots, 8$ ) is the standardized discriminant function coefficient, and  $Z_j$  ( $j = 1, 2, \dots, 8$ ) is the standardized score of the  $j^{th}$  explanatory variable.

The unstandardized and standardized discriminant function coefficients are presented in Table 4.10. These coefficients would be used like beta coefficients in multiple regressions. The discriminant function coefficients are partial coefficients, reflecting the unique contribution of each predictor variable to the classification of the households into their respective food security status.

#### 4.5.2.1 Unstandardized Discriminant Function Coefficients

The unstandardized coefficients are used to construct the actual prediction equation which can be used to classify new households. We can develop a discriminant function using the unstandardized coefficients given in table 4.10. The discriminant function score for the  $i^{th}$  ( $i = 1, 2, \dots, 828$ ) household is given as:

$$D_i = 0.238 - 0.245(HHSZ)_i + 0.319(GEND)_i + 0.057(EDUC)_i - 0.123(DIST)_i + 0.360(FRTLZR)_i + 0.131(TLU)_i + 0.351(OXEN)_i + 1.668(FLSZ)_i$$

Prediction of group membership of a household can be performed by using the discriminant score function given above. For instance, take a household with HHSZ=9, GEND=1(male), EDUC=0, DIST=12 km, FRTLZR=1(user), TLU=4, OXEN=0 and FLSZ=0.88 hectare, then the discriminant score of this household is -0.772. Since -0.772 is closer to the code 0 than the code 1; the household would be predicted food insecure household.

Interpretation of the two-group discriminant coefficients closely follows the logic of multiple regressions. Hence, the unstandardized discriminant function coefficients of the predictors included in the model are interpreted as follows:

Holding all other predictors constant the discriminant score decreases by 0.245 for one person increase in HHSZ. Consequently, the larger HHSZ the household has, the more likely the household will be food insecure. Holding all other factors constant, the value of discriminant score is expected to be higher by 0.319 for the male headed households than for the female headed households. Hence, male headed households are more likely to be food secure as compared to the female headed households. As EDUC increases by one year, the discriminant score increases by 0.057. Keeping the effect of other predictors constant, the higher the educational level of the head, the more likely the household will be food secure. As DIST increases by one kilometer, the discriminant score decreases by 0.123. Holding all other predictors constant, the farther the household reside from the agricultural input sources, the more likely the household will be food insecure. Holding all other factors constant, the discriminant score is expected to be higher by 0.360 for the fertilizer users than those who do not use fertilizer. As a result, households who use fertilizer are more likely to be food secure as compared with households who do not use fertilizer. As TLU increases by one unit, the discriminant score increases by 0.131. Holding the other predictors constant; the larger number of livestock the household own, the more likely the household will be food secure. Holding the other predictors constant; as OXEN increases by one ox, the discriminant score increases by 0.351. The greater number of oxen the household has, the more likely is the household to be food secure. Keeping the other variables constant the discriminant score increases by 1.668 for a hectare increase in farm land size. Hence, the larger farm land size the household has, the more likely the household will be food secure.

#### **4.5.2.2 Standardized Discriminant Function Coefficients**

The standardized discriminant function coefficients, also termed the standardized canonical discriminant function coefficients, are used to compare the relative importance of the predictor variables, much as beta weights are used in regression. Importance is assessed relative to the model being analyzed. Addition or deletion of variables in the model can change discriminant coefficients markedly. We get standardized coefficients for each variable in the discriminant function, and they can be interpreted as the larger (in absolute value) the standardized coefficient, the greater is the contribution of the respective variable to the discrimination between groups. Hence, based on the results given in Table 4.10, the contribution of the predictor variables to the discrimination of the two household groups from more important to the less important are ranked as: distance in km to

input sources, household size, farm land size, number of oxen, number of livestock in TLU, use of fertilizer, gender of head of household and educational level of head of household.

Table 4.10: Discriminant Function Coefficients

| Variables                                       | Unstandardized Coefficients | Standardized coefficients |
|---|-----------------------------|---------------------------|
| Household Size ( <b>HHSZ</b> )                  | -0.245                      | -0.535                    |
| Gender of head of household ( <b>GEND</b> )     | 0.319                       | 0.146                     |
| Level of Education of head of HH( <b>EDUC</b> ) | 0.057                       | 0.132                     |
| Distance in KM to input sources ( <b>DIST</b> ) | -0.123                      | -0.679                    |
| Use of Fertilizer ( <b>FRTLZR</b> )             | 0.360                       | 0.163                     |
| Livestock ownership ( <b>TLU</b> )              | 0.131                       | 0.297                     |
| Number of oxen ( <b>OXEN</b> )                  | 0.351                       | 0.336                     |
| Farm land size ( <b>FLSZ</b> )                  | 1.668                       | 0.366                     |
| (Constant)                                      | 0.238                       |                           |

The discriminant score for the  $i^{th}$  ( $i = 1, 2, \dots, 828$ ) household using the standardized coefficients from Table 4.10 is given as:

$$D_i = -0.535(Z_{HHSZ})_i + 0.146(Z_{GEND})_i + 0.132(Z_{EDUC})_i - 0.679(Z_{DIST})_i + 0.167(Z_{FRTLZR})_i + 0.297(Z_{TLU})_i + 0.336(Z_{OXEN})_i + 0.366(Z_{FLSZ})_i \quad \text{where}$$

$Z_j$  ( $j = 1, 2, \dots, 8$ ) is the standardized score of the  $j^{th}$  variable.

The results show that distance to input sources is the most important variable in discriminating the two groups as compared to the other variables included in this study. Agricultural inputs, such as seeds, fertilizers, chemicals, pesticides and other agricultural tools are very important to enhance agricultural production of the farmer, and consequently they help to ensure household food security. According to our result, distance to the sources of the inputs such as market centers, agricultural extension service provider centers, stores of fertilizer and improved seed and other

source centers has negative impact on the status of household food security. That is, households who reside far-away from the source centers may not have access to the important agricultural inputs timely and, hence, their over all agricultural output may decline which in turn causes the household to be food insecure.

In this study, household size was found to be the second most important predictor variable in discriminating the two household groups. From the results we see that the higher the household size, the more likely is the household to be food insecure. This is because the majority of farmers residing in the study area are small scale semi-subsistence producers with limited land holding size and limited non-agricultural activities. In areas like the rural part of Tigray region, different researchers (e.g, Kidane, 2005) have arrived to the conclusion that family size tends to exert more pressure on consumption than the labour it contributes to production. Hence, based on the results of the data taken from the sampled rural households of the Tigray region, this study verified that household size has a negative impact on household food security.

The third most important predictor variable in discriminating rural households based on their food security status is farm land size. According to the results presented in Table 4.10, the larger the farm land size, the more likely is the household to be food secure. The reason behind this is that the household can either cultivate the land to obtain more production or may rent it to people in short of cultivable land so as to generate more income to the household. Given equal family size and other predictors, households who own larger farm land size per capita are more likely to be food secured as compared to those who have smaller farm land size in Tigray region.

As compared to the predictor variables included in this study, the contribution of number of oxen owned by the farmer is ranked forth in discriminating food secure households from their counterparts. Exceptional to other livestock, oxen ownership is very crucial for the household because it is the source of traction power. According to the report on the livelihood profile of the region conducted in 2007 by the regional government, most of the poor residing in rural areas of the region do not have oxen. As a result, they rent-out their land primarily because they do not have drought power to utilize their entire land holding timely. The report also shows that the very poor at times exchange two days human labor for one day's oxen labor or they exchange crop

residues for cattle feed for one oxen. The results of this study, like similar other studies, proved that oxen ownership is one of the prominent factors for determining the status of household food security in the rural areas of the region. That is, households who own larger numbers of oxen are more probable to be food secure than those with lesser or no oxen.

Another important predictor variable is number of livestock ownership in TLU. This study revealed that households who own greater number of tropical livestock unit are given higher expectation to be categorized in the food secure households than their counterparts. This is because livestock are used as a source of income and food. Livestock are considered as valuable assets; they can be sold in the event of a hazard when other income sources are unavailable and households need access to relatively significant income. Livestock products such as milk, butter, meat, egg etc. are used directly for food or they can be sold to meet regular household expenses. Therefore, under the same conditions, households with greater number of livestock are more likely to be food secure as compared to the households who own lesser number of livestock.

Use of fertilizer is also found to be an important predictor variable in discriminating rural households into the two food security status groups. The result in table 4.10 depicts that fertilizer users are better to be food secure relative to the households who do not use fertilizer. According to the study conducted by Sosina (2004), adoption of fertilizer was found one of the major important strategies to increase crop production and to improve food security in Tigray region. Another similar study by Kidane (2005) on the “role of fertilizer in agricultural productivity” found that fertilization of farmland can boost agricultural production and influence the food security status of a household positively. These studies and other literatures imply that the likelihood of food security increases with farmers’ utilization of fertilizers. Consistent with previous similar studies, our results in this study show that fertilizer users are more advantaged in food securing strive than their counterparts in Tigray region

The second least important predictor variable as compared to the other variables included in this study is gender of household head. Although it has less contribution relative to the other variables, gender has a positive impact in discriminating the households into food secure and food insecure groups. From the standardized discriminant function coefficients portrayed in Table 4.10, we see

that male headed households are more likely to be food secure than female headed households in the study area. This is mainly because of differences between male and female heads to participate in non-farm activities that help to generate income (Adane, 2008). A likelihood profile conducted by the regional government in 2007 shows that the main constraints to food security in the female headed households are lack of labor availability and lack of livelihood assets such as land and oxen. Even when they have enough number of oxen and farmland size, female headed households are still more likely to be food insecure than male headed households because women can not plow their land as men can do timely. As a result, they need a man labor to plow their land in exchange of other expensive things such as cash, oxen labor, crop, straw for cattle feed etc., which consequently reduce their income and their status of food security.

The contribution of educational level of the household head in discriminating the rural households is the least as compared to the other predictor variables included in this study. However, educational level of the head was found important in determining household food security status. That is, based on the findings it can be observed that the higher the educational level of the head, the more likely is the household to be food secure. This is because literate farmers can read instructions and easily understand how to use fertilizers, pesticides, animal drugs etc. and they can easily implement new agricultural technologies. As a result, an educated household head is better adopter of new technologies such as rain water harvesting, use of fertilizer, use of improved seed and so on, which in turn increase the production of the household and level of food security status.

#### **4.5.4 Structure Matrix**

Structure matrix, also called structure coefficients or structure correlations or discriminant loadings, are the correlations between a given independent variable and the discriminant scores associated with the discriminant function. They are used to show how closely a variable is related to the function in the discriminant analysis. The correlations serve like factor loadings of predictor variables on the discriminant function in factor analysis. Note that in our case i.e., for two-group discriminant analysis, the structure coefficients show the order of importance of the predictor variables by total correlation, whereas the standardized discriminant coefficients (given in Table 4.10) show the order of importance by partial (unique) contribution.

Table 4.11: Structure matrix

| Variable                                 | Function 1 |
|--|------------|
| Distance in KM to input sources          | -0.625*    |
| Farm land size                           | 0.383*     |
| Livestock ownership in TLU               | 0.340*     |
| Number of oxen                           | 0.326*     |
| Household Size                           | -0.312*    |
| Child dependency ratio of household(a)   | -0.200     |
| Level of Education of head of household  | 0.151      |
| Use of Fertilizer                        | 0.132      |
| Gender of head of household              | 0.115      |
| Use of Agricultural extension service(a) | 0.046      |
| Average monthly expenditure (a)          | -0.030     |
| Use of improved seed(a)                  | -0.016     |
| Age of head of household(a)              | -0.001     |

(a). this variable is not in the analysis

\*. Largest absolute correlation in the analysis

The structure matrix given in Table 4.11 helps us to identify the largest (in absolute value) correlations associated with the discriminant function considered in this study. The variables are ordered by absolute size of correlation with the discriminant function. Based on this table, one can see that the predictors distance to input sources, farm land size, number of livestock, number of oxen, and household size load most heavily on the discriminant function. Educational level of the head, use of fertilizer and gender of the household head also load higher on the discriminant function as compared to the remaining predictor variables.

In general the higher the absolute value of the coefficient, the greater the discriminatory power of the predictor variable on the dependent variable. Hence, distance to input sources has the highest correlation with the discriminant scores, followed by farm land size, livestock ownership, number of oxen, household size and so on.

### 4.5.5 Group Centroids

One way to determine the degree of separation between the two groups is to compute the mean discriminant scores for each group. These means are called group centroids which are given in Table 4.12 below. From this table we see that the means are well apart (-0.941 & 0.689) showing the discriminant function is clearly discriminating the two household groups based on their status of food security.

Table 4.12: Functions at group centroids

|                         | Function |
|-------------------------|----------|
| HH Food Security Status | 1        |
| Food insecure           | -.941    |
| Food Secure             | .689     |

Unstandardized canonical discriminant functions evaluated at group means

The centroids presented in the above table are also used to establish the cutting point for classifying households. Since the group sizes are unequal in this case, the optimal cutting point is the weighted average of the two values, which is approximately zero. Therefore, households with discriminant score above the cutting point (0) are classified as “food secure”, while those households with discriminant score below zero are classified as “food insecure”.

We can use a cutting score to sort the households into either group based on their discriminant scores. Box-Whisker plot of the distributions of discriminant scores for food secure and food insecure households with the cutting score set at zero is given in Figure C5 ( Annex C).

The Boxplot show that almost all discriminant scores of the food insecure households lay below the cutting point while discriminant scores of food secure households lay above the cutting point. This shows that the discriminant function was efficient in predicting actual group membership of the households.

## 4.6 Classification results

Another major purpose to which discriminant analysis is applied in this study is the issue of predictive classification of households. Once a model has been finalized and the discriminant function has been derived, how well can we predict to which group a particular household belongs?

Table 4.13 given below tells us that the prior probabilities are 0.423 for group 1 (food insecure group) and 0.577 for group 2 (food secure group). These priors are taken into account to which group a household belongs. Ceteris paribus, one should be a little more likely to predict membership in group 2 than in group 1.

Table 4.13: Prior Probabilities for Groups

| HH Food Security Status | Prior | Cases Used in Analysis |          |
|-------------------------|-------|------------------------|----------|
|                         |       | Unweighted             | Weighted |
| Food insecure           | .423  | 350                    | 350.000  |
| Food Secure             | .577  | 478                    | 478.000  |
| Total                   | 1.000 | 828                    | 828.000  |

To assign cases into groups, a classification function is developed for each group. The classification functions are not to be confused with the discriminant functions. The classification functions can be used to determine to which group each household most likely belongs. There are as many classification functions as there are groups. Each function allows us to compute classification scores for each case for each group, by applying the formula:

$$C_k = w_0 + w_{k1}X_1 + w_{k2}X_2 + \dots + w_{kp}X_p$$

In this formula,  $C_k$  is the resultant classification score for the  $k^{th}$  ( $k = 1, 2$ ) group;  $w_{kj}$  ( $k = 1, 2$  &  $j = 1, 2, \dots, 8$ ) is the weight for the  $j^{th}$  variable in the computation of the classification score for the  $k^{th}$  group;  $w_0$  is a constant,  $X_j$  is the observed value for the respective case for the  $j^{th}$  variable.

Data for each household are inserted into each classification equation and the household is assigned to the group for which it has the highest classification score. A score on the function for group k ( $C_k$ ) is found by multiplying the raw score on each predictor ( $X$ ) by its associated classification function coefficients, summing overall predictors, and adding a constant  $w_0$ .

The classification function coefficients are given in Table 4.14 below. From the given results, we can develop classification functions for food secure and food insecure household groups separately.

Table 4.14: Classification Function Coefficients

| Variables                                 | HH Food Security Status |             |
|---|-------------------------|-------------|
|   | Food insecure           | Food Secure |
| Household Size ( $X_1$ )                  | 0.980                   | 0.581       |
| Gender of head of household ( $X_2$ )     | 1.251                   | 1.771       |
| Education level of head ( $X_3$ )         | 0.191                   | 0.285       |
| Distance in KM to input sources ( $X_4$ ) | 0.429                   | 0.229       |
| Use of Fertilizer ( $X_5$ )               | 1.912                   | 2.498       |
| Livestock ownership ( $X_6$ )             | 0.290                   | 0.503       |
| Number of oxen ( $X_7$ )                  | -0.288                  | 0.285       |
| Farm land size ( $X_8$ )                  | 7.058                   | 9.779       |
| (Constant)                                | -9.390                  | -8.485      |

Fisher's linear discriminant functions

**Food insecure group (1):** the classification function of this group is given as:

$$C_1 = -9.390 + 0.980X_1 + 1.251X_2 + 0.191X_3 + 0.429X_4 + 1.912X_5 + 0.290X_6 - 0.288X_7 + 7.058X_8$$

Larger household size and long distance from the input source centers contribute more strongly to classification of a household in this group. As the number of livestock in TLU, number of oxen, educational level of the head and farm land size per capita of the household decrease, the household is highly probable to be member of the food insecure households. The same is true for female headed and for households who do not use fertilizer.

**Food secure group (2):** the classification function of this group is given as

$$C_2 = -8.485 + 0.581X_1 + 1.771X_2 + 0.285X_3 + 0.229X_4 + 2.498X_5 + 0.503X_6 + 0.285X_7 + 9.779X_8$$

The probability of classification of a household into the food secure group increases as farm land size, number of livestock, level of education of the head and number of oxen of the household increase. Households with smaller household size and nearer to the agricultural input source centers are more likely to be assigned in this group. Moreover, male headed and fertilizer user households are much more likely to be classified in the food secure group.

The summary classification results given in Table 4.15 shows that the discriminant function was successful in predicting group membership for 398 (83.3%) out of 478 households who were actually members of group 2 and 262 (74.9%) out of 350 households who were actually members of group 1. Overall 79.7% of all surveyed households are correctly classified. This observed classification is significant ( $\chi^2 = 281.53, p < 0.0001$ ). Thus, observed classification is significantly different from expected chance classification. That is, the hit ratio (79.7%) of the actual classification is better than the expected chance (57.7%) classification. Note that by the maximum chance criterion, one could classify 57.7% of the households correctly.

Table 4.15: Classification Results

|          |       |                         | Predicted Group Membership |             | Total |
|----------|-------|-------------------------|----------------------------|-------------|-------|
|          |       |                         | Food insecure              | Food Secure |       |
| Original | Count | HH Food Security Status |                            |             |       |
|          |       | Food insecure           | 262                        | 88          | 350   |
|          |       | Food Secure             | 80                         | 398         | 478   |
|          | %     | Food insecure           | 74.9                       | 25.1        | 100.0 |
|          |       | Food Secure             | 16.7                       | 83.3        | 100.0 |

The misclassification rate also can be determined based on the classification summary results given in Table 4.15 above. From this table we found that the overall apparent error rate is 0.203. This error rate is an estimate of the probability that our classification function based on the present sample will misclassify a future observation. Based on the Rencher's (1995) rule, an error rate less than 0.30 is a more realistic estimate of what the classification functions can do with future

samples. Hence, our classification functions are more effective in classifying new households into their respective group.

Besides to the above classification results, the “casewise statistics” results obtained by SPSS are presented in ANNEX D (for only 20 households). This table lists the actual group and the predicted group based on largest posterior probabilities, prior probability, the Mahalanobis distance of the case to the group centroid and the discriminant functional scores for individual households. The actual group column gives the group membership of each household based on the food security status while the predicted group column gives us the most probable group of classification of the particular household. If a household is correctly classified we see the same values in both columns while when the household is misclassified the SPSS output displays asterisks (\*) beside the actual group number. When we see the results for the 20 households in ANNEX D the 5<sup>th</sup>, 8<sup>th</sup>, 9<sup>th</sup>, 14<sup>th</sup> and 15<sup>th</sup> households were misclassified.

In general, the discriminant analysis employed in this study was effective to the extent that 79.7% of original grouped households are correctly classified by use of selected determinants of food security in the rural areas of Tigray region.

# CHAPTER FIVE

## CONCLUSIONS AND RECOMMENDATIONS

### 5.1 Conclusions

The major objective of this study was to identify the determinants of food security among the rural households of the Tigray National Regional State. As a result, this study found that household food security in the study region was determined by eight key factors. However, the author believes that this is not a complete study to come up with solid solution to address the food security situation in the region under this study. This is because the range of factors and elements that affect food security are complex and multifaceted in nature and not easy to comprehend. Therefore, effort has been made in this study to see the impact of some demographic and socioeconomic factors on household food security.

In the study region 42.7% and 57.3% of the households were found food insecure and food secure, respectively. The figures show that the proportion of food secure households is higher than the food insecure households in the year during which the data was collected. However, this result might have been changed if the data had been collected in another year with deficient rainfall.

According to descriptive statistics of the sample households, the averages of some variables such as distance to input source and household size were found higher with food insecure households than the food secure households. On the other hand, the food secure households have relatively greater averages on the farm land size, educational level of the head, number of livestock and number of oxen than food insecure households. In addition, it was found that the male headed and fertilizer user households were better food secure as compared to their counterpart households.

The food security related factors studied through the discriminant analysis revealed that factors such as household size, gender of the household head, educational level of the head, distance to input sources, use of fertilizer, livestock ownership, number of oxen and farm land size were found the major contributors to the discrimination of the two group households and to the prediction of actual group membership of a household. Moreover, the predictors were ranked based on their importance to the discrimination of the two household groups. Accordingly, distance to input

sources was ranked first followed by household size, farm land size, number of oxen, livestock ownership, use of fertilizer, gender and educational level of the household head.

In this study, distance to the input sources was found the most important factor in aggravating food insecurity to the households. This is because farmers residing in the remote rural areas have problems regarding access to markets, agricultural extension service centers and other centers where agricultural inputs may be available. As a result, households who dwelled near to the main market centers have shown better food security status than those households located far as they are exposed to information distortion as well as transportation problem. Similarly, it was found out that large family size has high influence in worsening the food insecurity status of households. Our conclusion is that households with greater household size are more likely to be food insecure as compared with households with smaller household size.

Other principal determinants of household food security in the study region are number of livestock owned and amount of land to be cultivated. Land holding size was found one of the important factors in ensuring food security to the households. Farmers with greater farm land size showed better food security status than the less endowed households. Having large farm land size is not only essential to produce enough crops but also is a determinant factor for farmers to use new technologies such as fertilizers, improved seeds and so on. But the land holding in the region has been declining due to an increase in population. Similarly, households who own large number of livestock have better food security situation than those who have lesser. Livestock enables the households to be food secure either through the income earned or by direct consumption. But oxen hold a special place which relates to the amount of land farmers can cultivate. If a farmer owns a pair of oxen he can plough his own land and also rent in additional land from poorer neighbours without oxen in return for a 50% share of the harvest. Even owning one ox is a major advantage in securing food if it can be paired with that of a neighbour in similar circumstances. However, livestock holdings are limited by the lack of common grazing areas, and the consequent reliance, especially for cattle, on crop residues and collected/bought straw and grasses which only the better off can usually afford.

Furthermore, use of fertilizer, gender and educational level of the household head were found the crucial factors in determining the status of household food security in the study region. Use of fertilizer was found an important factor in enhancing crop production, which in turn improves household's food security status. Fertilizer using households in comparison with non-fertilizer users' households were better in food security status of the household. It can be concluded that fertilizer utilization boosts the overall harvest of the farm households, which in turn improves food security. On the other hand, the status of household food security was better in male headed households than female headed households. This leads to the conclusion that female headed households are more disadvantaged in food security than their male counterparts. Finally, it was realized that educational level of the household heads enables to improve the food security status of rural households. Education helps to change an attitude of the head, which in turn enables to adopt new agricultural technologies such as using fertilizer, using improved seed, pesticides, and storages and so on.

In general, with reference to a base group of food insecure households we conclude that an increase in land holding size, increase in oxen ownership, decrease in household size, decrease in distance to input sources, increase in fertilizer use, increase in educational level of household head, increase in livestock ownership and being male headed household increase the likelihood of a household to be classified into the group of food secure households in the study area.

## **5.2 Recommendations**

As rural part of Tigray region is constantly facing food insecurity and famines, there is a need for integrating famine relief and prevention strategies at the regional level with the overall development strategy. The strategy should aim at self-sufficiency at the local level and food security at the household level by incorporating the following recommendations.

- The longer the distance that the farmers travel from their home to the agricultural input sources, the more food insecure they are likely to be. Thus there is a need to formulate intervention strategies by the local and federal governments to work jointly in order to alleviate the transportation problems and build a corporate institute that can supply agricultural inputs and provide information about the market situation. A policy which provides adequately

trained and equipped extension workers for disseminating improved agricultural inputs to the remote rural areas should also be designed.

- It should be noted that household size is known to be one of the leading causes of food insecurity in the study area. This implies that policy measures directed towards the provision of better family planning to reduce household size should be given adequate attention and priority by the federal and regional governments. Education that encompasses all aspects of training and which brings about attitudinal changes targeting at reducing fertility level is important for rural households in the study area.
- With increase in population size of the region, land is becoming very limited. It is therefore important that the regional government should design a policy which enables farmers to increase their landholding size. A possible intervention to increase the size of landholding of farmers in the short run may be resettling volunteer households to other parts of the region, where better land resources available.
- The livestock sub-sector is very important in assuring food security in the study region. Hence, this sector has to be enhanced through the provision of common grazing land, better husbandry and management system, and better veterinary facilities. Development of livestock also helps in crop production by raising the traction power and manure. From all livestock resources oxen are strategic asset especially for farming households; since they serve as a source of traction in the rural households. Therefore, concerned bodies should support the poor farmers by providing access to draught power.
- Increasing the productivity of major cereal crops is feasible through the provision of cultivable land, provision of education to farmers, and application of important agricultural inputs such as fertilizer. Fertilizer utilization happens to supplement the food security of households. This implies that there is a need to increase fertilizer users among the poor households through the provision of credit service at a lower interest rate and opening fertilizer distribution centers at the remote rural areas.

- The study has provided evidence that gender of head of a household play a key role in determining food security status of households. Thus, gender-sensitive food insecurity alleviation policies that enhance endowments of female-headed households should be a key ingredient of food insecurity reduction strategy.
- Based on the study, households with educated heads are better in food security status than households with non-educated heads in the study region. Therefore, it is recommended that the regional and federal governments should provide access to education for farmers. In the short-term, informal education could be effective, especially when targeted at farmers who have had limited formal educational opportunities.
- Finally, we recommend for further studies to be conducted on the area of food security by considering detail and accurate information on various variables including political, climatic and weather (rainfall and temperature), topography, natural disasters, ecological conditions and other factors that affect food security. It is also recommended to conduct a study that compares status of food security in rural households with urban households.

Generally, food insecurity is a multifaceted concept, which cannot be treated in isolation from other causes of poverty. Therefore efforts geared towards achieving food security should be addressed to the areas of human and infrastructure development.

## References

- Adane, T. (2008). Correlates of Household Food Insecurity in Daro-Luba Woreda West Hararge Zone of Oromia National Regional State, Ethiopia; Thesis Paper, Addis Ababa University.
- Alex, D. (2003). HIV/AIDS and Food Security in Africa, A Report for DFID.  
[http://www.livelihoods.org/info/info\\_guidancesheets.htm](http://www.livelihoods.org/info/info_guidancesheets.htm)
- Amaza, P.S., Umeh, J.C., Helsen, J. and Adejobi, A.O. (2006). Determinants and Measurement of Food Insecurity in Nigeria: Some Empirical Policy Guide. Contributed poster prepared for presentation at the International Association of Agricultural Economics Conference, Gold Coast, Australia, August 12-18, 2006.
- Anderson, T.W. (1984), *An Introduction to Multivariate Analysis*, New York: Wiley series.
- Babatunde, R.O., Owotoki, G.M. and Heidhues, F. (2007). Vulnerability and Food Insecurity Differentials Among Male and Female-headed Farming Households in Nigeria: *Pakistan Journal of Social Sciences* 4(3): 414-418.
- Barret, C. (2002). *Food Security and Food Assistance Programs*. In B.L. Gardner and G.C. Rausser (eds). *Hand book of Agricultural Economics* (volume 2). Elsevier Science, Amsterdam
- Berhanu, A. (2004). The food Security Role of Agriculture in Ethiopia. Ethiopian Economic Policy Research Institute. *eJADE*. 1:138-153.
- Boussard, J. and Benoit, D. *et al.* (2005). Food Security and Agricultural Development in Sub-Saharan Africa: Building a Case for More Support. Background Document for FAO, Rome.
- Clay, E. (1997). Food security: A status review of The Literature, Research Report ESCOR No.R5911. London: Overseas Development Administration.
- Clover, J. (2003). Food Security in Sub-Saharan Africa. African security review.  
Tigray Regional Government (2007). Livelihood Profile of Tigray Region ;Central Mixed Crop Livelihood Zone, Tigray, Ethiopia.
- CSA (2005). The 2005 National Statistics, Addis Ababa, Ethiopia.
- CSA (2007). Household, Income, Consumption, and Expenditure (HICE) Survey 2004/5, Analytical Report, Volume I: Statistical Bulletin 394.
- CSA (2007). Welfare Monitoring (WM) survey 2004/5, Statistical Report on Basic Population Characteristics in Ethiopia.

- CSA (2008). Summary and Statistical Report of the 2007 population and Housing census, Federal Democratic Republic of Ethiopia population Census Commission.
- Debebe, H. and Tesfaye, Z. (1995). Food Security: Conceptual Framework and Situation Analysis, Addis Ababa, Ethiopia.
- Degefa Tolossa (2002). Household Seasonal Food Insecurity in Oromia Zone, Ethiopia.
- Diana P.(2007). Food Security in Development countries: A systematic Perspective, Medellin-Colombia.
- Devereux, S. (2000). Food Insecurity in Ethiopia. A discussion paper for DFID <http://www.fivims.net/documents/Final%20Paper5.pdf>
- Devereux, S.(2006). Distinguishing Between Chronic and Transitory Food Insecurity in Emergency Needs Assessments, Institute of Development Studies, Rome Italy.
- Devereux, S. and Maxwell, S.(2003). Food Insecurity in Sub-Saharan Africa, University of Natal press piemartzburg, South Africa.
- Driba, G.(1995). Economy at the Crossroad-famine and Food security in Rural Ethiopia. Care International in Ethiopia, Addis Ababa.
- Dulla Etana ( 2007). Demographic Daynamics and Sustainable Rural Livelihood in Jimma Arijo District, Eastern Wellega of Oromia, Ethiopia; Thesis paper Addis Ababa University.
- EEA (2005): Ethiopian Economic Association. A research paper on Land Tenure and Agricultural development in Ethiopia. Addis Ababa, Ethiopia
- EHNRI (1998): Ethiopian Health and Nutrition Research Institute. Food Composition Table for use in Ethiopia, part IV.
- FAO World Food Summit (1996): *The State of Food Insecurity in the World*, Rome.
- FAO (2002): Food and Agricultural Organization. The State of Food Insecurity: Report 2002 Food and Agricultural Organization, Rome, Italy.
- FAO (2005): Food and Agricultural Organization. Assessment of the World Food Security Situation, committee on world food security. Thirty-first sessions, 23-26 May 2005, Rome
- FARM-Africa (1998). FARM Africa Community Oriented Rural Development project,Tigray, Ethiopia
- Fasil Gebrekiros (2005). Enough with Famine in Ethiopia, Commercial printing enterprise, Addis Ababa.

- Fasil, F. and Mukti, P. (2007). Determinants of Household Supply of Labor in Food –for-Work Program in Tigray; Ethiopia. Southern Illinois University, London.
- Government of Ethiopia (2003). The New Coalition for Food Security in Ethiopia.
- FDRE FSS (1996): Federal Democratic Republic of Ethiopia. *Food Security Strategy*
- Fisher, R.A. (1936). The Utilization of Multiple Measures in Taxonomic Problem. *Annals of Eugenics* 7:179-188.
- Fitsum Hagos, Pender, J. and Nega, G. (2002). Land Deradation and Strategies for Sustainable Land Management in The Ethiopian Highland: Tigray Region Socio-Economic and policy Research Working paper, International Livestock Research Institute.
- Frank, R., Nancy, M., Bruce, C. et al. (1999). Food Security Indicators and Framework for Use in the Monitoring and Evaluation of Food Aid Programs; Food and Nutrition Technique Assistance
- Getachew Driba (1995). Economy at the Cross Road, Famine and Food Security In Rural Ethiopia: Commercial printing enterprise.
- GOA (2008): United States Government Accountability Office. Food Insecurity in Sub-Saharan Africa, Washington, Dc 20548.
- Gujarati, N.D. (1988). *Basic Econometrics*, New York, Economic Series.
- Haddinot (1999). Choosing outcome indicators of household food security. IFPRI, Washington, D.C., U.S.A.
- Haile, K, Alemu G., and Kudhlande,G. (2005). Causes of Household Food Insecurity in Koredegaga Peasant Association, Oromia Zone, Ethiopia: working paper.
- Haswell, M.(1953). Economics of Agriculture in Savannah Village: Colonial Research Study, No.8.
- IFPRI (2002): International Food Policy Research Institute. Proceedings from sustainable Food Security for all by 2020. An International Conference (4-6 September 2001, Bonn). IFPRI, Washington, D.C., U.S.A
- Johnson, R.A. and Wichern D.W. (2002): *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.
- Karzanowski W.J. (1980). Mixture of continuous and categorical variables in Discriminant Analysis, *Biometrics*, 36, 493-499.
- Keshav, L. (2006). Household food Security in Rural Areas of Nepal: Relationship between Socio Economic Characteristics and Food Security Status, Hiroshima University, Japan.

- Kidane, H., Alemu, Z., & Kundhlande G. (2005). Causes of household food insecurity in Koredegaga peasant Association, Ormia Zone, Ethiopia. *Agrekon* 44(3): 543-560.
- Kidane (2006). Responding to Food Security Challenges Through an Integrated Watershed Approach: The Tigray Experience, Tigray, Ethiopia.
- Maja, P., Mateja, B., and Sandra, T. (2004). Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study. *Metodoloski*, Vol.1, No.1, 143-161.
- Majda, B.S. (1999). Food Security for the Food-Insecure: new challenges and renewed commitments, CSD Women's Caucus Position paper for CSD-8, 2000.
- Markos Ezra (1997). Demographic Response to Ecological Deradation and Food Insecurity, Drought Prone Areas in the Northern Ethiopia: PDOD publications.
- Maxwell, S., Belache, D., and Lirensa, A. (1994). The disincentive effect of food-for-work on labor supply and agricultural intensification in Ethiopia. *Journal of Agricultural Economics* 45(3):351-359.
- Maxwell, S. and Frankenberger, T. (1992). Household Food Security: Concepts, Indicators, Measurements, IFAD and UNICEF, Rome, Italy.
- MEDaC (1999): Ministry of Economic Development and Cooperation. Poverty Situation in Ethiopia. Government of Ethiopia, Welfare Monitoring unit, Addis Ababa.
- MoFED (2002): Ministry of Finance and Economic Development. Sustainable Development and Poverty Reduction Program, Ethiopia.
- Mwanki, A. (2005). Achieving Food Security in Africa-Challenges and Issues. Cornell University, USA.
- Ogbu, J. (1973). Seasonal Hunger in Tropical Africa as Actual phenomenon. *Africa* No.43.
- Osman Ali and Tesfahuna Fenta (2003). Challenges and Properties of Food Security in Ethiopia: Proceeding of Food Security Conference in 2003, Professional Association of Joint Secretariat.
- Ramakrishna, G. and Assefa, D. (2002). An Empirical Analysis of Food Insecurity in Ethiopia: The case of North Wollo. *African development*, Vol.27:127-143.
- REST (1995): Relief Society of Tigray. Farming systems, Resource management and households coping strategies in Northern Ethiopia
- Rencher, A. C. (1992). *Canonical Discriminant Functional Analysis*. *American statistics* 46, 217-225
- Rencher, A.C. (1995). *Methods of Multivariate Analysis*, New York: Wiley series

- Salih, A. (1994). Food Security in East and Southern Africa; World Institute for Development Economics Research, *Nordic Journal of African Studies* 13(1):3-27.
- Seber, G.A.F. (1994). *Multivariate observations*. New York: John Wiley & sons.
- Sen, A. (1981). *Poverty and Famine: An Essay on Entitlement and Deprivation*; Oxford, Clarendon press.
- Shiferaw, F., Kilmer, R., and Galdwin, C. (2003). Determinants of Food Security in Southern Ethiopia- a Selected Paper Presented at the 2003 American Agricultural Economics Association Meeting in Montreal, Canada, *Agricultural Economics* 33: 351-363.
- Sosina, B. and Stein H. (2007). Impact of Food-For-Work in Tigray, Ethiopia.
- Tabachnick, G.B. and Fidell, S.L. (1996). *Using Multivariate Statistics*. Harper Collins College Publishers: New York.
- Tabachnick, G.B. and S.L. Fidell (2001). *Using Multivariate Statistics*, Fourth Ed. Boston: Allyn and Bacon.
- Tabachnick, G.B. and Fidel, S.L. (2007). *Using Multivariate Statistics*, New York: Pearson Int.Ed.
- Toulmin, C. (1996). Access to Food, Dry Season Strategies and Household Size amongst The Bambara of Central Mali. *IDS Bulletin* 17, No.3, 58-66
- Tweeten, L. (1997). Food Security. In L.G. Tweeten and D.G. McClelland (eds).
- UNDP (1992). The Human Development Index- going beyond income.  
[Http://www.hdrstas.undp.org/country-fact-sheet/cty-fs-ETH.htm](http://www.hdrstas.undp.org/country-fact-sheet/cty-fs-ETH.htm).
- USAID (2008): United States Agency International Development. Madagascar Food Security Development programming Framework, From the American People, Madagascar
- Wald, A. (1944). On a Statistical Problem Arising in the Classification of an Individual into one of Two Groups. *Ann.Math. Stat*, 15, 145-162.
- Webb, P., Braun, J., Yohannes, Y., (1992). Famine in Ethiopia: Policy Implications of Coping Failure at National and Household levels. *Research Report 92*, IFPRI
- World Bank (1986). *Poverty and hunger: Issues and Options for Food Security in Developing Countries*. The International Bank for Reconstruction and Development, The World Bank, Washington, D.C.

- World Bank (2007). Accelerating Equitable Growth, Ethiopia Country Economic No.38662-Momerandum, Poverty Reduction and Economic Management Unit Africa Region.
- Yared Amare (1999). Household Resources, Strategies and Food Security in Ethiopia: A Case Study of Amhara Households in Wagda, North Shewa Monograph of Society and Anthropology Vol.1. The Department of Sociology and Social Administration and Addis Ababa press.
- Yared Amare *et al.* (2000). Food Security and Resource Access: Final Report on Community Assessment in South Wollo and Oromia Zones of Amhara Region, Ethiopia.
- Yared, A. (2000). Livelihood Strategies and food security policy in Ethiopia; Addis Ababa, Ethioipa.

## Annexes

### ANNEX A: Tables of conversion factors

Table A1: Conversion factor used to estimate Tropical livestock Unit (TLU)

| Animal     | TLU   |
|------------|-------|
| Cattle     | 1     |
| Sheep/Goat | 0.15  |
| Horse      | 1     |
| Mule       | 1.15  |
| Donkey     | 0.65  |
| Camel      | 1.45  |
| Poultry    | 0.005 |

Source: G. Ramakrishna and Assefa Demeke, 2002.

Table A2: Conversion factor used to calculate adult equivalence scales

| Age groups<br>(years) | Male | Female |
|-----------------------|------|--------|
| 0-2                   | 0.40 | 0.40   |
| 3-4                   | 0.48 | 0.48   |
| 5-6                   | 0.56 | 0.56   |
| 7-8                   | 0.64 | 0.64   |
| 9-10                  | 0.76 | 0.76   |
| 11-12                 | 0.80 | 0.88   |
| 13-14                 | 1.00 | 1.00   |
| 15-18                 | 1.20 | 1.00   |
| 19-59                 | 1.00 | 0.88   |
| 60+                   | 0.88 | 0.72   |

Source: World Bank (1986)

Table A3: Caloric content of the food consumed in the study area

| <b>Group No.</b> | <b>Food items</b>         | <b>Food energy in Kcal per 100 grams portion</b> |
|------------------|---------------------------|--|
| <b>1</b>         | <b>Cereals</b>            |  |
|                  | <b>Maize</b>              |  |
|                  | White porridge            | 154.70   |
|                  | White bread               | 223.40   |
|                  | Injera                    | 153.00   |
|                  | Whole roasted             | 88.10  |
|                  | White kitaa               | 223.4  |
|                  | <b>wheat</b>              |  |
|                  | Bread                     | 222.00   |
|                  | Kitaa                     | 222.00   |
|                  | <b>Teff</b>               |  |
|                  | Injera                    | 358.80   |
|                  | Porridge                  | 165.40   |
| <b>2</b>         | <b>Vegetables</b>         |  |
|                  | Onion                     | 71.30  |
|                  | Cabbage                   | 40.10  |
|                  | Tomato                    | 30.70  |
|                  | Green pepper              | 46.50  |
| <b>3</b>         | <b>Livestock products</b> |  |
|                  | Milk                      | 73.70  |
|                  | Meat                      | 212.30   |
|                  | Egg                       | 295.10   |
|                  | Cheese                    | 132.40   |
|                  | Butter                    | 736.40   |
| <b>4</b>         | <b>Others</b>             |  |
|                  | oil                       | 896.40   |

Source: Ethiopian Health and Nutrition Research Institute (EHNRI, 1998).

## ANNEX B: Test of Assumptions

Table B1: Collinearity information of predictor variables

| Variable          | Eigenvalue | Condition Index | VIF  | 1/VIF    |
|-------------------|------------|-----------------|------|----------|
| <b>HHSZ</b>       | 0.82377    | 3.50460         | 1.74 | 0.574204 |
| <b>GEND</b>       | 0.59064    | 4.13884         | 1.10 | 0.907782 |
| <b>AGE</b>        | 0.53410    | 4.35240         | 1.12 | 0.896517 |
| <b>EDUC</b>       | 0.39506    | 5.06070         | 1.04 | 0.965698 |
| <b>DIST</b>       | 0.29061    | 5.90044         | 1.04 | 0.95960  |
| <b>AGREXT</b>     | 0.27186    | 6.10056         | 1.22 | 0.821874 |
| <b>FERTILIZER</b> | 0.26028    | 6.23478         | 1.16 | 0.860363 |
| <b>TLU</b>        | 0.22683    | 6.67864         | 1.20 | 0.834899 |
| <b>CDR</b>        | 0.17934    | 7.51103         | 1.48 | 0.674301 |
| <b>OXEN</b>       | 0.13141    | 8.77464         | 1.19 | 0.842495 |
| <b>FARMLSZ</b>    | 0.08980    | 10.61483        | 1.08 | 0.929098 |
| <b>EXPEND</b>     | 0.06241    | 12.73303        | 1.31 | 0.762109 |
| <b>SEED</b>       | 0.02615    | 19.67182        | 1.21 | 0.825570 |

Table B2: Coefficients of Skewness and Kurtosis for continuous variables

| Variable       | Skewness  | Kurtosis   |
|----------------|-----------|------------|
| <b>HHSZ</b>    | 0.1735735 | -0.1170958 |
| <b>AGE</b>     | 0.5385584 | -0.4688594 |
| <b>EDUC</b>    | 2.4095972 | 5.3926535  |
| <b>DIST</b>    | 0.4917509 | 0.2610774  |
| <b>TLU</b>     | 0.4797607 | -0.3709337 |
| <b>CDR</b>     | 1.0722361 | 2.4569356  |
| <b>OXEN</b>    | 0.7970587 | -0.4259038 |
| <b>FARMLSZ</b> | 0.6044737 | -0.2873338 |
| <b>EXPEND</b>  | 0.5030361 | -0.2958311 |

## ANNEX C: Normal Q-Q plots, Chi-square plots and Box plot

Figure C1. Normal Q-Q Plot of Discriminant Scores for all HHs

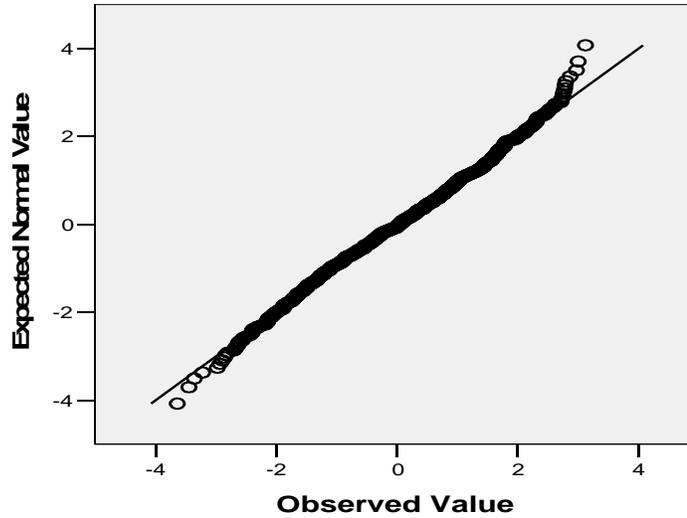


Figure C2: Chi-square plot for all household

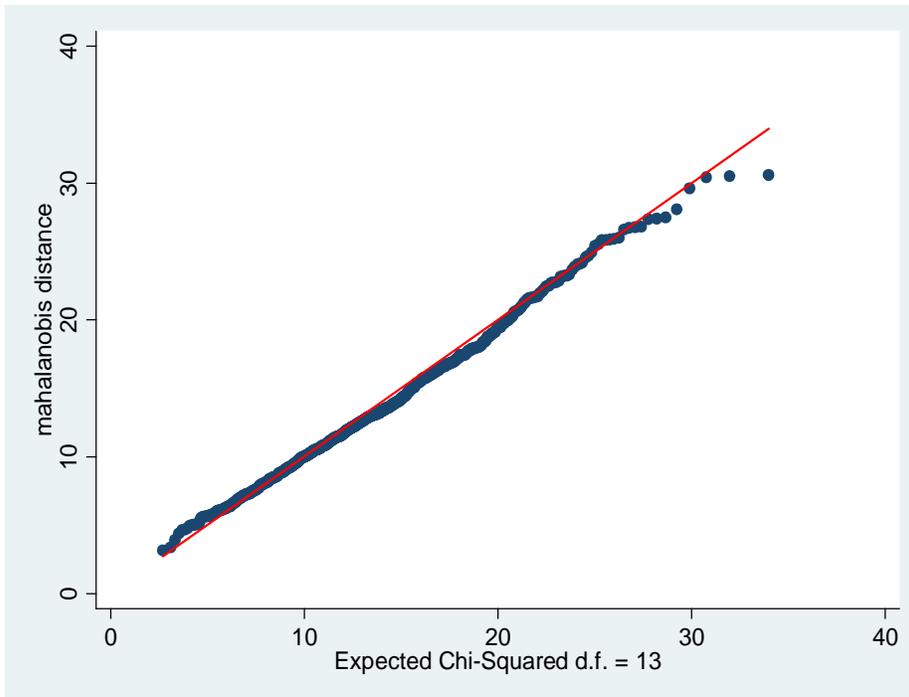


Figure C3: Chi-square plot for food secure households

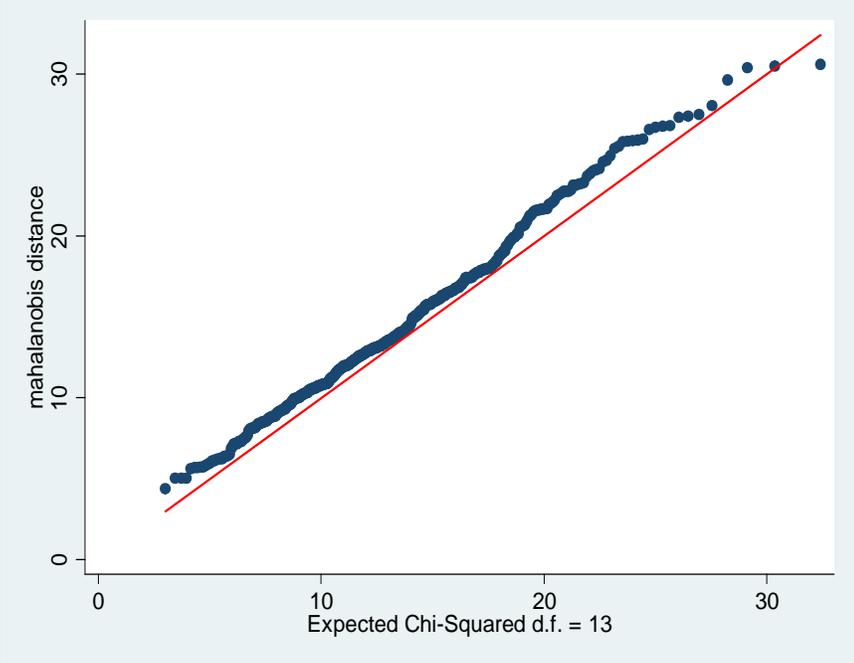


Figure C4: Chi-square plot for food insecure households

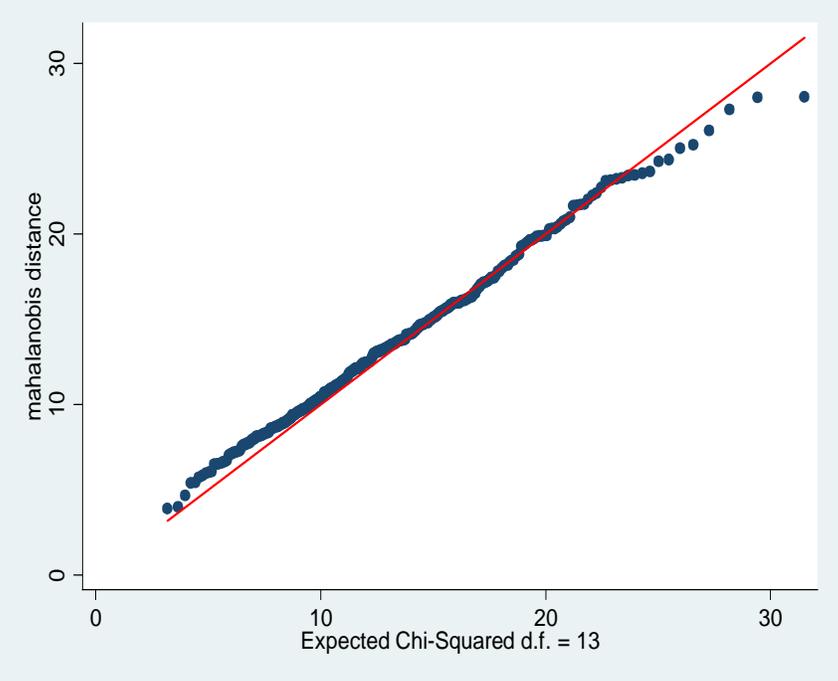
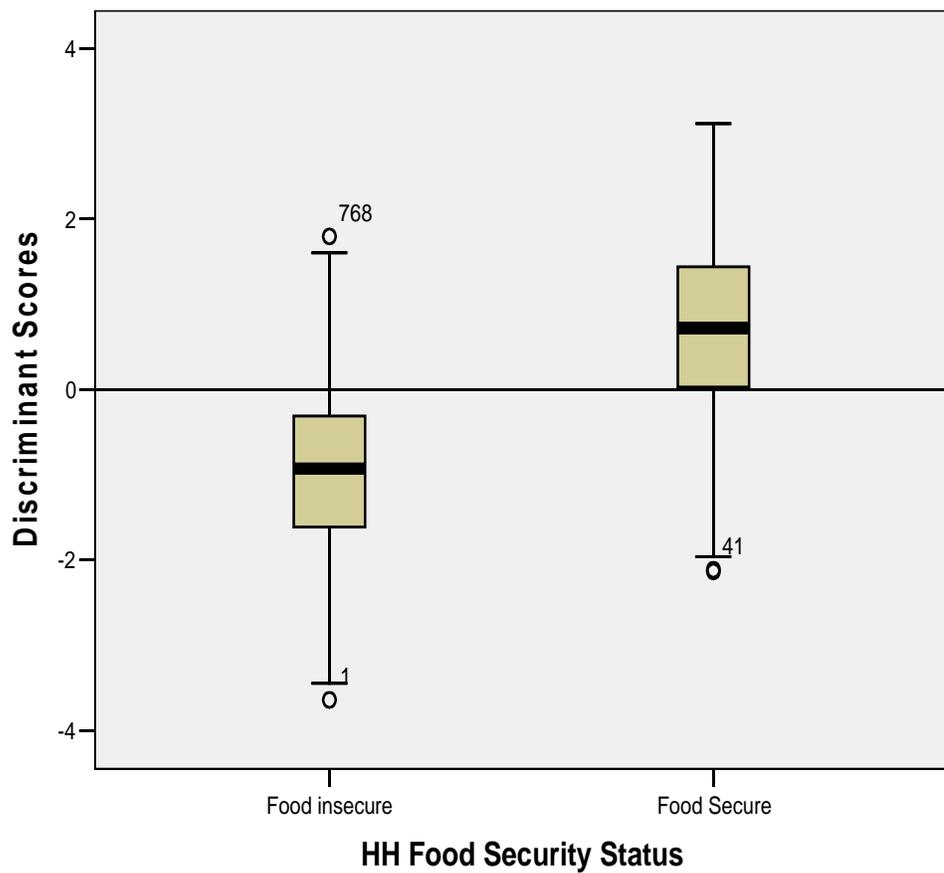


Figure C5: Boxplot of discriminant scores



## ANNEX D: Casewise Statistics

Table D5: casewise Statistics

| Case Number | Actual Group | Predicted Group | Highest Group |    |              |  | Second Highest Group |              |  |        | Discriminant Scores<br>Function 1 |
|-------------|--------------|-----------------|---------------|----|--------------|--|----------------------|--------------|--|--------|-----------------------------------|
|             |              |                 | P(D>d   G=g)  |    | P(G=g   D=d) | Squared Mahalanobis Distance to Centroid | Group                | P(G=g   D=d) | Squared Mahalanobis Distance to Centroid |        |                                   |
|             |              |                 | p             | df |              |  |                      |              |  |        |                                   |
| Original 1  | 0            | 0               | .580          | 1  | .529         | .306                                     | 1                    | .471         | 1.161                                    | -.388  |                                   |
| 2           | 1            | 1               | .789          | 1  | .769         | .072                                     | 0                    | .231         | 1.858                                    | .422   |                                   |
| 3           | 0            | 0               | .268          | 1  | .944         | 1.228                                    | 1                    | .056         | 7.501                                    | -2.050 |                                   |
| 4           | 0            | 0               | .368          | 1  | .923         | .809                                     | 1                    | .077         | 6.401                                    | -1.841 |                                   |
| 5           | 0            | 1**             | .359          | 1  | .958         | .843                                     | 0                    | .042         | 6.495                                    | 1.607  |                                   |
| 6           | 1            | 1               | .849          | 1  | .791         | .036                                     | 0                    | .209         | 2.072                                    | .498   |                                   |
| 7           | 1            | 1               | .058          | 1  | .991         | 3.605                                    | 0                    | .009         | 12.455                                   | 2.588  |                                   |
| 8           | 0            | 1**             | .428          | 1  | .586         | .628                                     | 0                    | .414         | .702                                     | -.103  |                                   |
| 9           | 0            | 1**             | .583          | 1  | .678         | .301                                     | 0                    | .322         | 1.170                                    | .140   |                                   |
| 10          | 1            | 1               | .423          | 1  | .583         | .643                                     | 0                    | .417         | .687                                     | -.113  |                                   |
| 11          | 1            | 1               | .968          | 1  | .846         | .002                                     | 0                    | .154         | 2.792                                    | .730   |                                   |
| 12          | 1            | 1               | .053          | 1  | .992         | 3.749                                    | 0                    | .008         | 12.721                                   | 2.625  |                                   |
| 13          | 0            | 0               | .474          | 1  | .899         | .512                                     | 1                    | .101         | 5.504                                    | -1.657 |                                   |
| 14          | 0            | 1**             | .485          | 1  | .623         | .487                                     | 0                    | .377         | .870                                     | -.009  |                                   |
| 15          | 1            | 0**             | .644          | 1  | .565         | .214                                     | 1                    | .435         | 1.364                                    | -.479  |                                   |
| 16          | 1            | 1               | .327          | 1  | .511         | .959                                     | 0                    | .489         | .424                                     | -.290  |                                   |
| 17          | 1            | 1               | .333          | 1  | .516         | .936                                     | 0                    | .484         | .440                                     | -.278  |                                   |
| 18          | 1            | 1               | .616          | 1  | .695         | .251                                     | 0                    | .305         | 1.275                                    | .188   |                                   |
| 19          | 1            | 1               | .357          | 1  | .959         | .848                                     | 0                    | .041         | 6.511                                    | 1.610  |                                   |
| 20          | 0            | 0               | .042          | 1  | .987         | 4.138                                    | 1                    | .013         | 13.430                                   | -2.975 |                                   |

\*\* . Misclassified case

## Declaration

I, the undersigned, declare that the thesis is my original work, has not been presented for degree in any other university and that all sources of material used for the thesis have been duly acknowledged.

Declare by

Name-----

Signature-----

Date-----

This thesis has been submitted for examination with my approval as a University advisor.

Name-----

Signature-----

Date-----