بسم الله الرحمن الرحيم

SUDAN UNIVERSITY OF SCIENCE AND TECHNOLOGY

COLLEGE OF GRADUATE STUDIES

USING LOGISTIC REGRESSION MODEL TO ANALYSE AND COMPARE FACTORS AFFECTING POVERTY

For the National Household Budget Poverty Survey Data (2014-2015)

إستخدام نموذج الانحدار اللوجستي لتحليل ومقارنة العوامل المؤثرة على الفقر

لبيانات المسح القومي لميزانية الأسرة والفقر (2014-2015)

Thesis Submitted in Fulfillment of the Requirements for the Doctor of Philosophy Degree in statistics

By

Ismail Abaker Adam Bandh

Supervisor

Dr. Hamza Ibrahim Hamza

Co-Supervisor

Dr. Altaiyb Omer Ahmed

November 2018

CHAPTER ONE

INTRODUCTION

1.1 Preface

1.2 The Problem of the Study

1.3 The importance of the Study

1.4 The Objectives of the Study

1.5 The Hypotheses of the Study

1.6 The Methodology of the Study

1.7 The Data Sources

1.8 The Limited of the Study

1.9 Review of Previous Studies

1.10 Comparison between this Study and Previous Studies

1.11 Organization of the Research

INTRODUCTION

1.1 Preface:

Once stated that “No society can surely be flourishing and happy, of which the greater part of the members are poor and miserable”. Poverty, as a social phenomenon, can be described most simply as a state of scarcity that can be felt in many aspects of an individual’s life (Smith, 1776). While traditionally spoken of as a lack of income, nowadays poverty is being redefined to include non-monetary aspects in which scarcity of resources also stresses the lives of those in poverty like lack of freedom or political rights, lack of opportunities, lack of access to valuable resources like education or healthcare, lack of safe environments, and social discrimination (United Nations, 2005). It is only logical that fighting to eradicate this phenomenon, and preventing people from ever experiencing it, is at the core of the Millennium Development Goals (MDG) agreed by at least 183 countries. According to these goals, that set the tone for the development agenda since 1990, the signing countries agreed to fight poverty and by 2015 reduce in 50% the number of people living under the international line of extreme poverty, set at 1.25 US dollars a day.

The results of this fight against poverty have been diverse. While countries like China have accounted for most of the world’s decrease in poverty headcounts, countries in Sub-Saharan Africa still have more than half the worlds extremely poor. Latin America is a region of great disparities; whereas the poverty rates are not as alarming as in other regions they are still not par to those in Europe or other developed regions. And although extreme poverty might be low, income inequality is one of the most worrying issues, being the region with the largest disparities in the world.

A minority of the world's population (17%) consume most of the world's resources (80%), leaving almost 6 billion people to live on the remaining 20%., as a result, billions of people are living without the very basic necessities of life - food, water, housing and sanitation (Rich, 2013). Specifically, 1.2 billion (20%) of the world population now lives on less than $1/day, another 1.8 billion (30%) lives on less than $2/day, 800 million go to bed hungry every day, and 30,000 - 60,000 die each day from hunger alone (World Bank, 2013).

In fact, poverty is the most risk, complicated and most measured issues, and despite its widespread and increasing in all of the developing countries, in Sudan it became a frightening matter.

Poverty remains widespread; almost half of Sudanese live below the national poverty line, with 75% of the poorest living in rural areas. Poverty varies widely across regions; the lowest poverty incidence is in Khartoum State (26%), and the highest is in the Darfur region (62.7%). One-quarter of the population has no access to health facilities, the employment rate of the population is only 68%, and the highest unemployment is found among youth age 15-24 years (over 20%) living (CBS, NHBPS, 2009). Such that, it is observable that the problem of poverty became affecting the exhaustion of diverse and available resources in these country, specifically the human, physical and environmental resources, which require redirecting the effort towards poverty alleviation by reading its specificities and understanding its patterns and causes, in reference to the social, economic and cultural situations in which the poor people.

Generally, poverty indicates a status of absolute deprivation of one or more of dimensions of welfare for an individual, such as education, the lack of accessing healthcare facilities, decreased human capital, the insufficiency of accommodation infrastructure, malnutrition, and lacking some of commodities and services, the inability of expressing political opinions, and the like, and each one of these dimensions deserves a separated consideration (Mustafa, 2011).

The main concern of this study, however, is not to define, nor measure poverty but to study the factors that act as determinants of such phenomenon and analyze their behavior through a survey. Are there any particular characteristics that make people more prone to be in a situation of poverty or that are highly correlated with it? Using national household budget and poverty survey was undertaken by Central Bureau of Statistics (2014/2015), a logistic regression model is constructed that explains the probability of poverty occurring in the household as a function of a set of socio-demographic characteristics of the household head. The marginal effect of each correlate is then calculated and its evolution is plotted and analyzed throughout the period studied.

The aim is to gain a clear enough picture of the evolution of the effects of the determinants of household poverty in Sudan.

1.2 The Problem of the Study

Poverty is a global phenomenon and has proven difficult to resolve and remained determinedly high in developing a country. We have exacerbated the problem of poverty in Sudan particularly, which requires the studying this phenomenon, and its helps in the formulation of economic and social policies to address them. Poverty as a multidimensional concept includes monetary and nonmonetary characteristics. The study examines which household and personal characteristics of the household head as the determinants of poverty status in Sudan. Poverty in Sudan is mainly a result of a very complex history and cannot be understood without reference to the Government's policy and that reflected the impact of socioeconomic and demography variables on poorness. The problem of the study mainly on how can we come out with the main effective factors that affected the poverty household status and showing the relationships between the response variable and the explanatory variables?

1.3 The Importance of the Study

The importance of the study stems from its attempt to learn the concept of Poverty and its measurement methods and to know the standard methodologies that followed by studies of poverty profiles in Sudan to provide a comprehensive understanding of the phenomenon of poverty. We need to know the most important determinants affecting household poverty status. During of these studies, we know the relationship between the socioeconomic determinants and household and the reflection of this relationship with poverty.

1.4 The Objectives of the Study:

These study objectives at investigating the severity of the poverty phenomenon and the extent of its spread in Sudan. The main objective of the study is estimation and analysis of the factors affecting poverty household status, by using logistic regression analysis, more specifically, it at:

1. To review the poverty concepts and assess the situation of determinant factors and house poverty status in Sudan.

2. To examine the poverty characteristics and identify the demography and social and economic factors which influenced on household poverty in Sudan.
3. To identify the most important demography, social and economic factors that influence poverty household in Sudan.
4. To establish and determine the relationship between demographic and socio-economic characteristics and poverty among households in Sudan.

1.5 The Hypotheses of the Study

The research will be based on analytical and interpretive hypotheses, with the hope that they will be including all the elements of the subject of the study, and these hypotheses are represented as follows:

1. There is a significant relationship between the demography factors set and the poverty.
2. There is a significant relationship between the social factors set and the poverty.
3. There is a significant relationship between the economic factors set and the poverty.
4. There is a significant relationship between the demography, socio-economic factors, and the poverty.

1.6 The Methodology of the Study

The research was based on the descriptive, and analytically approaches, to achieve the objectives of the study; the descriptive side used in the use of frequency tables, percentages of data, and cross-tabulations was run to establish the relationship between the dependent and independent variables. The analytical aspect of constructing the binary logistic regression model and checking the quality of the data, through books, references, researches, and papers related to the theme. In-depth analysis using logistic regression model has been applied, to identify the significant variables, and describing the relationship between a response variable and explanatory variables of poverty households, and it's determined the importance of independent variables and their impact on poverty, also use Quality statistics tests, model evaluation, and data estimation was used with the help of the Statistical Packages for Social Sciences SPSS and STATA for analysis and research findings.

The study attempts to construct a logistic regression model considering indicators like household expenditures per capita are considered dependent variables. The set of independent variables, that are included in the model of the determinants of poverty in Sudan some important socioeconomic and demography factors. Socio-economic variables are very often categorical, rather than interval scale. In many cases research focuses on models where the dependent variable is categorical. We could not carry out a multiple linear regression as many of the assumptions of this technique will not be met. Instead, we would carry out a logistic regression analysis. Multiple linear regression models are suitable when the key response is a quantitative measurement variable, while logistic regression models are applicable when the key response variable is binary, i.e., the response takes only two possible values (e.g., yes/no, poor/not poor).

1.7 The Data Sources

This research depends entirely on quantitative and qualitative raw data file that will be collected from National Household Budget and Poverty Survey undertook by Central Bureau of Statistics (2015), which is the most recent, available at the time this study is written. It is the fourth in the series of such surveys undertaken by Central Bureau of Statistics. The first (NHBPS) was conducted in 1967; the second one was in 1978; the third in 2009 and currently, there is national household budget and poverty survey (2015). Also, we depend on NBHS 2009 data and from other sours like researches, books and papers.

1.8 Limits of the Study

The time limitation: the research includes the period of National Household Budget Poverty Survey (2014/2015). The place limitation: includes eighteen States of Sudan.

1.9 Review of Previous Studies

1. Mohammed, 2017. (Measurement and Determinants of Urban Poverty in Case of Southern Nations, Nationalities, and Peoples' Region (SNNPR), Ethiopia). This study used survey data collected by Southern nations, nationalities and peoples' region (SNNPR) bureau of finance and economic development (BFED) for 5,015 urban households. The major purposes of the research are measuring urban poverty and identifying the determinants via employing logistic regression. Accordingly, for the year 2015 poverty incidence, gap and severity were equal to 18.02%, 5.25%, and 2.31% respectively for the urban SNNPR. Urban food poverty measured using the above indexes leaves relatively larger figures. In the last five years, the region's urban poverty reduces remarkably except for food poverty severity which rose by 17.24%. The use of logistic regression to identify the determinants of urban poverty end up with marital status, family size, total dependency, education level, saving habit, and source of energy was found to be statistically significant variables. Hence, pre and post marriage orientations to reduce divorce and input support for windowed limiting family size and in turn dependency using short and long-term solutions, and supplying social and physical infrastructure such as education, financial institutions and power are viable options to reduce urban poverty in the region.

2. Tuyen, 2015. (Socio-Economic Determinants of Household Income among Ethnic Minorities in the North-West Mountains, Vietnam) This paper investigates both commune and household determinants of household income among ethnic minorities in the North-West Mountains the poorest region of Vietnam. The findings show that the vast majority of the sample households heavily depend on agricultural activities. Factors affecting household income per capita are examined using multiple regression models and the findings confirm the important role of education, non-farm employment and fixed assets in improving household income. In addition, some commune variables such as the presence of the means of transportation, post offices and, non-farm job opportunities are found to have an increasing impact on household income.

3. Paola & Duclos, 2015. This paper assesses (multidimensional poverty in Sudan and South Sudan).

We use the National Baseline household Surveys (NBHS) of 2009 to measure poverty incidence in education, consumption, access to public assets and possession of private assets across these two countries. Our findings show regional and sub-population differences in the one-dimensional and multidimensional poverty status of people in Sudan and South Sudan. Poverty in Sudan is generally less severe than in South Sudan, with a pattern showing (i) lesser One-dimensional incidence of poverty; (ii) lower multidimensional poverty indices and prevalence, but similar breadth, in Sudan than in South Sudan, both for adults and children. This pattern also points towards Khartoum and Western Equatorial as the states with the least poverty, and Northern Darfur, and Warap as the states with the greatest poverty, both for adults and children, in Sudan and South Sudan, respectively. The policy intended at reducing poverty in each of the two countries should recognize the poverty profile differences across age groups, geographical areas, and dimensions.

4. Garza, 2015. (Determinants of poverty in the Mexican states of the US-Mexico border) This study examines the determinants or correlates of poverty in the Mexican states bordering with the United States. The data used in the study come from the 2008 National Survey of Income and Expenditures of Households. A logistic regression model was estimated to determine which variables might be important in explaining poverty in this region. It was found that the variables which are positively correlated with the probability of being poor are: living in Coahuila, Tamaulipas or Chihuahua, size of the household, being an ambulatory worker or working in an agricultural occupation, and being a manufacturing, transportation, sales, domestic service or support worker. Variables that are negatively correlated with the probability of being poor are living in Baja California, the education level of the household head and his/her age.
Gender of the household head and household location were not statistically significant in the logistic regression analysis.

5. Yusyf, 2015. Study about, (Determinants of Rural Poverty in Tanzania) this study aimed at assessing the determinants of poverty in Mkinga district in rural Tanzania. The ordinal regression model was used to model events of observing scores of livelihood status in the area of study.

The study revealed that nearly 93% of respondents in the area were poor. Gender, size of land the household owns, the size of the farm used in farming, Household size and the dependency ratio were found to be related to poverty, hence influencing poverty in the area of study. While the government is responsible for providing proper infrastructural settings, this research recommends that people especially women in this area should be empowered to have positive attitudes towards participating in economic activities using resources around them.

6. Ngunyi, 2015. (Multidimensional Analysis of the Determinants of Poverty Indicators in the Lake Victoria Basin (Kenya)), this study main objective is to examine the multidimensional aspects of poverty in one of Kenya's culturally diverse region of the Lake Victoria basin. The analysis using data collected by IUCEA researchers in 2007 and also the 2009 census on households in Kenya; the research findings indicate that poverty measures do overlap to capture a percentage of the sample as poor. The analysis shows that education, gender (being male), marital status, assets (livestock, water sources, and wall materials) and age of the head of the family have statistically positive effects on the likelihood of an individual falling into poverty.

7. Habyarimana, Zewotir, & Ramroop, 2015. (Analysis of demographic and health survey to measure poverty of household in Rwanda), this study used the principal component analysis PCA technique in order to create the asset index. Then the asset index was used to assess the socio-economic status SES of households. The methodology is applied and demonstrated using the household survey data in Rwanda. The Rwanda data analysis showed that the age of household head, education level of the household head, gender of the household head, place of residence, the province of household head and size of the household (number of household members) were the significant predictors of poverty of the household in Rwanda.

8. Farah, 2015. (Impact of Household and Demographic Characteristics on Poverty in Bangladesh), the main objective of this paper is to identify the factors that have a relative effect on poverty of the household. The principal component analysis was used to create an asset index which gave the Social Economic Status (SES) of each household.

The variables were tested as a univariate model to see the effect on SES. Finally, a logistic regression was estimated based on this data with the SES (that is poor and non-poor) as the dependent variable and a set of demographic variables as the explanatory variables. The results presented in this paper suggest that the DHS data can be used to determine the correlates of poverty. The results also suggest that demographic and household data can describe poverty. The probability of a household being poor depends on the ownership of assets and other household data. A closer look was then taken to identify whether the results were driven by the rural or urban property.

9. Balarabe, 2014. (Empirical Investigation of the Determinants of Poverty in Kano Metropolis, Nigeria**),** this study provides some explanations of the causes of poverty in Kano metropolis by investigating poverty determinants that are too often neglected in the literature and in policy debates. The study comprises six local governments in the state which includes Dala, Fagge, Gwale, Municipal, Nasarawa, and Tarauni. Primary data was collected using a questionnaire and interview from one hundred and twenty (120) residents selected in the study area. The data were analyzed using a probit regression analysis and the result showed that all coefficients of the explanatory variables have a positive relationship with poverty except that of education which has a negative relationship.

10. Spaho, 2014. (Determinants of Poverty in Albania), the aim of this study is to identify the determinants of poverty in Albania, at the household level using a questionnaire. The data were collected during November 2013, and direct interviews were conducted with 215 households living in the rural and urban area. Two regression models were estimated based on the collected data, a log-linear model with the logarithm of per capita monthly consumption as the dependent variable and a logistic model with poverty status as the dependent variable, and a set of economic and demographic variables as the explanatory variables. It was found that the variables that impacted the per capita consumption of the household and the poverty status of the household were household size and residence. Poverty alleviation efforts should be made to improve the social and demographic characteristics of the households since the number of the poor is increased in both urban and rural areas. To reduce poverty, great attention must be paid to the manufacturing sector, agriculture and tourism.

11. Makame, & Mzee, 2014. (Determinants of Poverty on Household Characteristics in Zanzibar): A Logistic Regression Model. The two succession of Zanzibar Household Budget Survey (ZHBS) in 2004/2005 and 2009/2010 use headcount to address poverty as the base of all analysis with several social and economic variables. This study attempts to use logistic regression to venture ratio of the probability of occurrence of poverty in Zanzibar with a social dimension. The study reveals that social demographic dimensions are important in explaining poverty and that the likelihood of poverty significant relates to household size, household head, and basic education (primary and secondary). Furthermore, the study exposes that all district in Pemba is at high risk of entering into poverty.

12. Xhafaj, & Nurja, 2014. (Determination of the key factors that influence poverty through econometric models**)** the purpose of this study is to determinate the most important factors that influence poverty through econometric models that are the logistic regression and the linear log regression. The data are obtained from the Living Standards Measurement Study (LSMS) for 2008 that includes 3600 households interviewed in Albania. The result of both econometric models confirms that the variables those are strongly connections with the expenditure of consumption per capita and with the economic status are: household size, the educational level and gender of the head of household, the zone. This study recommends a careful review on the reforms to be taken in relation to education in Albania.

13. Adetayo, 2014.(Analysis of farm households poverty status in Ogun states, Nigeria**),** this study examined the poverty status of farm households in Ogun State, Nigeria using a descriptive statistics, Foster, Greer and Thorbecke poverty (FGT) indices and Logit regression model.

The data used were generated from a survey involving 117 farm household's randomly selected using a multistage sampling technique. Results of the analysis revealed that majority (70.9%) of the farm households do not have access to potable water; they live in mud buildings while the common toilet facility was the bush.

The mean per capita consumption expenditure among the farm households was 9,103.85 with the FGT poverty incidence, poverty gap, and severity of poverty estimated to be 78.1%, 55.8% and 43.0% respectively. Poverty incidence was found to be higher among male-headed (60%) and farming (63.9%) households and those having over five members (66.1%). The logit regression further indicates that the likelihood of being poor was more with large households, non-educated farm households head and households without access to credit and other non-farm income.

14. Abd Razak, Ibrahim, 2014. (Determinants of poverty in Malaysia), this study it is important to summarize information on poverty in Malaysia and identify characteristics of the poor. Secondary data of 2007 Malaysia's Household Income Survey (HIS), focusing on lowest quintile household income group had been utilized and was extracted from the Economic Planning Unit, Prime Minister's Department Malaysia. Binary logistic regression was applied to identify the determinants of poverty. Based on the results of the logistic regression, this study discovered that age of head of households, household's size, number of income recipients, strata, the gender of head of households, marital status, education level and occupation of head of households are the factors contributing to poverty.

15. Zyka, & Bici, 2014. (Identifying Household Level Determinants of Poverty in Albania**)** Study about Identifying Household Level Determinates of Poverty in Albania Using Logistic Regression Model, the general goal of this paper is to analyze poverty in household per capita consumption as a monetary measurement and based in the data from Albania trying to identify probable determinants that influence in falling in a trap of poverty. Current literature suggests several ways of modeling the determinants of poverty. Usually, the regression analysis is used to check in the same time the influence of the different factors.

In this paper, binary logistic regression was estimated with economic status (poor-non poor) as the dependent variable and a set of characteristics of individual and household as independents variables. The logistic model used shows that the probability of being poor is found to be influenced mainly by education and status of employment of household head, the household composition and geographic divisions.

16. Khudri, & Chowdhury, 2013. (Evaluation of socio-economic status of households and identifying key determinants of poverty in Bangladesh**),** this study aims to evaluate living standards and socio-economic status of Bangladeshi households through constructing an asset index and identify key determinants of poverty in Bangladesh using the data extracted from 2007 Bangladesh Demographic and Health Survey (BDHS). The principal component analysis was applied to 72 dichotomous variables (owning the particular item or not) including ownership of durable goods, housing characteristics, and access to basic services to create the asset index. Ownership of land and dwelling made of cemented floor, roof, and wall indicated a positive impact on the socio-economic status of any household whereas the poor source of drinking water, sanitation facilities, and dwelling made of low-quality construction materials had a negative impact on the index. Using logistic regression model, a set of demographic variables such as division, type of place of residence, own land usable for agriculture, highest education level and employment status were identified as key determinants of poverty. The results also revealed that ownership of agricultural land, having higher education reduce the likelihood of being poor whereas rural and unemployed people were more prone to poverty.

17. Bahta, & Haile, 2013. (Determinants of Poverty of Zoba Maekel of Eritrea: A Household Level Analysis**),** this study for Determinants of Poverty of ZobaMaekel of Eritrea: A Household Level Analysis The descriptive result of mean per capita food expenditure (MPKFE) in ZobaMaekel of Eritrea found that all the households' heads are poor. The result of the Probit analysis shows that poverty status is strongly associated with almost all variables used. Education level, type of resident, size of land, number of the meal, remittance, access of credit from relatives, credit institutions, opinion to credit, rain-fed crop, irrigated crop, income from agriculture and income from -non-agriculture were found to be negatively associated with food self-sufficiency as a proxy of poverty. However, family number, number of children, children at school age and rent of land highly positively related to poverty.

For instance, higher levels of educational attainment will provide higher levels of welfare for the household. Education is not sufficient condition to escape from poverty. Remittance is a good indicator of poverty, showing strong family ties within Eritrean society, the fact that Eritrea does not have social security system it may help to pursue policies which foster cultural ties and family networks as part of poverty alleviating endeavor.

18. Sekhampu, 2013. (Determinants of poverty in a South African township), this study reported here used household-level data to analyze determinants of household poverty in a South Africa Township of Bophelong. A Logistic regression was estimated based on this data with the economic status (that is poor and non-poor) as the dependent variable and a set of demographic variables as the explanatory variables. The results show that household size, age and the employment status of the household head significantly explain the variations in the likelihood of being poor. The age and employment status of the household head reduces the probability of being poor, whereas household size is associated with an increased probability of being poor. The strongest predictor of poverty status is the employment status of the household head.

19. Gounder, 2012. (The determinants of household consumption and poverty in Fiji**),** this paper uses household survey data to model the determinants of household consumption and poverty in Fiji. A multivariate analysis is conducted to ascertain those household characteristics that correlate with household welfare and poverty. The ordinary least squares (OLS) estimation results suggest that higher levels of education, supporting agricultural growth policies in rural areas and reallocation of labour to the formal sector of the economy are likely to be effective in reducing poverty at the household level. The robustness of the results is checked by estimating a probit model. The probit estimates show the coefficients are robust to an alternative empirical approach.

20. Sinnathurai, & Březinová, 2012. This study about (Poverty Incidence and its determinants in the estate sector of Sri Lanka), the objective of this study is to find out and analyze the significant determinants of the incidence of poverty in the estate sector where the highest level of chronic poverty and unemployment exist.

The econometric model was fitted and estimated in this study. The Ordinary Least Square (OLS) regression analysis clearly indicates that variables such as industrial employment, education, access to market and infrastructure significantly and negatively affect the poverty incidence of the estate sector. Also, agricultural employment has a negative impact but not significant. Analysis with the Durbin–Watson stat confirms that, there is no autocorrelation between the variables. The results emphasize the need for adopting policies for regional infrastructural improvement as well as market and educational development in the plantation sector.

21. Abaker, & Salih, 2012. This study about, (income poverty and inequality in the service sector of Sudan), this research paper aims to address income poverty and inequality in the service sector of Sudan. Poverty and inequality indicators were computed using both primary and secondary data sources. P-alpha equation, Povstat, and Sims models were used for poverty measurement and simulation. Results showed that more than 82 percent of employees in the services sector living with poverty. Employees in health, education, transportation and security and justice were the poorest followed by the public sector, communication, and commerce respectively. The inequality measured Gini index was 43 percent with large inequality among commerce employees. The future prospect of poverty among employees in services, growth would slightly reduce poverty and inequality. The combined effect of growth and food prices increase would also reduce poverty and inequality. However, the decomposition of this effect into income and distribution effects, income effect would reduce poverty while the distributional effect would increase moderate poverty indicators and reduce food poverty indicators.

23. Gounder, 2011. This paper uses household survey data to model (the determinants of household poverty in Fiji). A multivariate empirical analysis is conducted to ascertain those household characteristics important in determining household welfare and poverty. The ordinary least squares (OLS) estimation results show that higher levels of education, supporting agricultural growth policies and reallocation of labour into the formal sector of the economy will prove effective in reducing poverty at the household level.

The robustness of the results is checked by estimating a probit model. The probit estimates show the coefficients are robust to an alternative empirical approach.

24. Sakuhuni, Chidoko, Dhoro, & Gwaindepi, 2011.This study investigates the empirical (economic determinants of poverty in Zimbabwe) using cross-section data for 2005. A regression model was estimated based on this data, with per capita consumption as the dependent variable and a set of economic and demographic variables as explanatory variables. Variables that are significant and positively correlated with per capita consumption thus negatively correlated to poverty are: age squared, gender (male), widow, maximum level of education, attaining primary education, employment in any sector except working in the informal sector, migration status, engaged in secondary business, number of sources of income, credit availability and land area cultivated. Variables significant and negatively correlated with per capita consumption and positively correlated to poverty are age and household size. Employment in the informal sector, days missed due to illness, and land ownership were some of the insignificant variables.

25. Ahmed, Roghim, & Saleh, 2011. (Poverty Determinants in South Sudan): This paper aimed to identify and analyze the main determinants of poverty in South Sudan prior to it secession from Sudan in 2011. Primary data were collected using a structured household questionnaire. A sample of 200 households was interviewed in Renk County. Multiple Regression analysis was used for estimating poverty determinants. The results of the determinants analyses indicated that secondary education, widow household heads, female household heads, government and private sector employees, petty traders, Gango, dysentery infection, the mixed source of water are the main poverty determinants in the urban area. While university education, married household heads, household size, female household heads, farmers, Gango, petty traders, total agricultural Original Research Article land, goats' ownership and numbers of chicken per households are the rural poverty determinants.

As spending on education, health, drinking water, and electricity are not only the responsibility of the households but also of the government.

26. Achia, Wangombe, & Khadioli, 2010. This study examines the determinants of poverty in Kenya. While most of the studies done on poverty determinants rely on the income, expenditure and consumption data, the data used in this study comes from the Demographic and Health Surveys, (DHS). The principal component analysis was used to create an asset index which gave the social economic status of each household. A Logistic regression was estimated based on this data with the SES (that is poor and non-poor) as the dependent variable and a set of demographic variables as the explanatory variables. The results presented in this paper suggest that the DHS data can be used to determine the correlates of poverty.

27. Brück, Danzer, Muravyev, & Weifphaar, 2008. This paper analyses the incidence, the severity and the determinants of household poverty in Ukraine during transition using two comparable surveys from 1996 and 2004. We measure poverty using income and consumption and contrast the effects of various poverty lines. Poverty in both periods follows some of the determinants commonly identified in the literature, including greater poverty among households with children and with less education. We also identify specific features of poverty in transition, including the relatively low importance of unemployment and the existence of poverty even among households with employment. Poverty determinants change over time in line with the experience of transition and restructuring.

28. Sikander, M.U., and Ahmed, M., 2008. (Household Determinants of Poverty in Punjab**),** this paper tries to model the various demographic and socio-economic determinants of poverty in Pakistan by using a logistic regression analysis. The results show that age, education, and gender of the household head significantly explain the variations in the likelihood of being poor. Moreover, households receiving remittances and holding agriculture land are more likely to exit from the poverty trap. The dependency ratio and larger family size positively affect the possibility of entering the poor household group.

The employment sector also significantly explains the cross-regional and geographical differences in the poverty determinants. The empirical results for the three mutually exclusive regions of the Rural, Other Urban and the Major Cities suggest considerations for the policymakers and provide poverty dynamics over these regionally differentiated localities.

29. Kamgnia, and Timnou, 2008. (The determinants of poverty in Cameroon), this study made use of data on the Cameroon household survey (ECAM III) collected by the national institute of statistics in 2007. Variables that explain household economic well-being were years of schooling of household head, the household head having a public or private sector job, the age of household head, gender, distance to the nearest hospital, distance to the nearest road, owning farmland, both urban and rural localities. From the regression results, the following variables contributed positively to household expenditure account; years of schooling of household head, access to credit, the household head having a public or private sector job and a number of migrants in the household. Variables that rather reduced household expenditure were the age of household head, owning farmland, male-headed households, and household head unemployed, distance to the nearest tarred road and distance to the nearest hospital. Based on these findings, the study advocates for government intervention with policies that encourages attainment of higher levels of education, employment, and rural development.

30. Halo, 2006. The paper about (human poverty index account). This study aimed at measuring the five basic dimensions contained in human poverty in Sudan based on the 1993 census data the study depended on finding a mean weighted arithmetic for indicators of human poverty in Sudan, the study found human poverty proportional extrusive with indicators. The study found that the percentage who die under the age of 40 to about 27.6% and illiteracy ratio between adult 49.9% and the proportion of deprived living necessities 27.5% and calculate these indicators led to the consequence of human poverty in Sudan is 35%.

31. Bogale, Hagedorn, & Korf, 2005. This paper investigates (the determinants of rural poverty in Ethiopia).
Our study is based on information gathered from a three-round survey of 149 rural households in three districts of Ethiopia during the 1999/2000 cropping season.

It reveals that nearly 40% of the sample households live below poverty line with an average poverty gap of 0.047. The binary logit estimates shed light on factors behind the persistence of poverty and indicates that rural poverty is strongly linked to entitlement failures understood as lack of household resource endowments to crucial assets such as land, human capital, and oxen. Our findings suggest that improved targeting devices can be a useful instrument in reducing poverty, in particular, to reach the poorest of the poor.

32. Geda et al., 2005. This study about (Determinants of poverty in Kenya), strategies aimed at poverty reduction need to identify factors that are strongly associated with poverty and that are amenable to medication by policy. This article uses household level data collected in 1994 to examine probable determinants of poverty status, employing both binomial and polychotomous logit models. The study shows that poverty status is strongly associated with the level of education, household size and, engagement in the agricultural activity, both in rural and urban areas. In general, those factors that are closely associated with overall poverty according to the binomial model are also important in the ordered-logit model, but they appear to be even more important in tackling extreme poverty.

33. Ministry of social welfare, 2004.Study about (definition and measurement of poverty and efforts to combat it at Khartoum). This study aimed at reaching a national definition of poverty in Sudan, are common human heritage Foundation and national privacy in dogma and culture and the degree of economic and social development and providing the database. The study reached the Islamic concept broader and more comprehensive and includes the necessities and needs improvement and everything related to the well-being and human dignity. And accommodate all measurement methods.

34. Justino, & Litchfield, 2003. (Poverty dynamics in rural Vietnam): This study identifies the transmission mechanisms of the impact of trade reforms on household poverty dynamics, based on data from a panel of rural Vietnamese households. Poverty dynamics are modeled using multinomial logistic regressions of poverty transition outcomes.

These models are shown to provide important insights into the behavior of poor households that cannot be explicitly derived from household consumption models. We find that changes in household poverty status in Vietnam are strongly correlated with price and employment changes induced by the trade reforms. These results are robust to shifts in the poverty line and changes in model specification.

35. Fofack, 2002. (The nature and dynamics of poverty determinants in Burkina**),** this study investigates the determinants and dynamics of poverty during the five-year growth period which followed the 1994 CFA franc devaluation in Burkina Faso. Results show that the nature and dynamics of poverty determinants are influenced by the spatial location of households and that the post-devaluation growth period did not significantly alter the pattern of poverty determinants. The most significant determinants of poverty over the growth period include the burden of age dependency, human and physical assets, household amenities and spatial location. Though consistently significant at the national level, the direction of the association between these determinants and welfare depends on the nature of the determinants. While the burden of age dependency is consistently negatively associated with welfare, asset ownership is positively associated; the probability of being poor declines with increasing share of household assets, and increases with the burden of age dependency. There are some variations at the regional level, however, illustrated by the difference in the scope of significance of these determinants. While the ratio of age dependency remains the most the significant determinant of rural poverty, its explanatory power reduces considerably in urban areas where its marginal effect on the probability of being poor is relatively low over the two reference periods, despite the significance of the probit coefficient and the relatively low asymptotic standard error.

36. Ismael, 2001. (Economic Growth, Resource Allocation and Poverty Reduction in Sudan (1990-2000))" The important findings this study is: Despite the good performance of Sudan economy in achieving high GDP growth rates there no reduction in poverty measured in terms income, access to education and access to health, the reason for this unequal distribution of income and resources.

Despite the increase in the number of schools, students and teachers in both primary and secondary level education; general education has become of poor quality. There no quality, this revealed by the high rates of drop – out in both rural and urban areas. The enrollment ratio in both primary and secondary schools is found to be low, which to reveals the inability to absorb all children in school age in basic education. Despite the increase in the number of hospitals and others health institutions, the mortality rate still remains high. In general life expectancy at birth also remains low and hospital and hospital beds (per 100.000 pop) are also still low.

1.10 Comparison between this Study and Previous Studies:

Previous studies at the local level that dealt with the issue of poverty are many, and all of them contribute to the resolution of poverty problem in different ways and various methodologies, most of the studies done on poverty determinants depend on the income and rarely used logistic regression model and national survey data in Sudan (see Minstry.2006, Ahamed, Roghim & Saleh. 2011and Abaker,Salih, 2012). We have problems with this procedure represented in data on household incomes are known to be less reliable than consumption data obtained from household expenditure surveys. But this study depended on expenditures of household's data to analyze the factors affecting poverty in Sudan by using logistic regression. However, there are other dimensions also not searched much, such as the demography variables which are a need for concentration.

At the international level, there are a huge number of researches and scientific papers which deal with the poverty and they depend on quantitative methodologies which dominate over the economic, social, developmental and demographic literature. Also, there are research papers dealt with the same issue and have been applied in developing countries, however sometimes the economic, social and cultural status of these countries are different from Sudan. Majority of that studies use expenditure data and logistic regression model (see Farah, 2015, Sekhampu, 2013 and Gounder. 2012). Furthermore, the previous studies dealing with different variables such as sex, age, education agriculture, income, and expenditure, but I use most of them in this study and add new variables like a place of resident, main sources of livelihood, dependency ratio and type of dwelling.

This study comes as a complementary to the other previous studies in the domain of the demography and the socio-economic factors affect on poverty and to formulate the required policies and programs to reduce its ratio by logistic regression.

1.11 Organization of the Research

It contains five chapters, the first chapter is the introduction includes the preface, and then outlines the problem statement, the importance of the study, research objectives, hypotheses, methodology, and review of previous studies and the organization of the study. Chapter two outlines poverty phenomenon concepts and explains the determinants of poverty. Chapter three logistic regressions and descriptive statistics Chapter four analyses of the data and interpretation. Chapter five is including the conclusions and recommendation and finally, references and appendix.

CHAPTER TWO

POVERTY

2.1 Preface

2.2 Conceptual of Poverty

2.3 Poverty in Sudan

2.4 The Poverty Line in Sudan

2.5 Poverty Measures in Sudan

2.6 Concept of Socio-economic Factors

POVERTY

2.1 Preface

Poverty is the main problem of the world in developing countries. The poor and unstable growth of GDP, unemployment, illiteracy and high age dependency ratio are common issues in the most of the developing countries such as Asia, Africa, and Latin America. The people of these countries have been afflicted by poverty and hunger over the long period.

According to (Sachs, & McArthur, 2005) more, than1.3billion people in the world are living on less than one dollar a day. 2.7 billion People in the world are living on less than two dollars a day. But it is the fact that approximately half of the world population are under poverty ridden condition in terms of international poverty line US $ 2. Eleven million children die every year. 114 million children are not able to achieve basic education and 584 million women are illiterate. In every year, Six million children die due to malnutrition. Every day 800 million people stay hungry in which 300 are children. 2.6 billion People of the world's population 40% are depriving of basic sanitation and one billion people are suffering to unsafe drinking water.

Over the 200 million people in Africa are trapped in the net of abject poverty. In Sub-Saharan Africa, the incidence of poverty is manifestly tremendous (Osinubi, 2005).On the average 45% to 50% of sub-Saharan Africans live below the poverty line in terms of $1.25 (World Bank,1997, Osinubi,2005). But in terms of $2, it is rather high (66.2%) in Sub-Saharan Africa. In West Africa, virtually all countries are classified as low-income countries by World Bank. In these countries, human poverty afflicts about half of the population (Ogwmike, 1998).

According to (Williams, 2004) in all these countries such as Asia, Sub- Saharan Africa and Latin America, chronic poverty is being transmitted to next generations because of unstable and poor economic growth, high population growth rate, and lack of education, severe unemployment, low paid wage and refusal of social and political freedom.

If we observe the natural resources of the world then we can examine that African countries and South American countries are probably the richest countries. But these are the world's poorer countries.
The countries that are poor in their natural resources such as England, Hong Kong, Japan, and Taiwan are prospering people of the world. One can argue that some countries emerged after colonialism. If there is such a thing then Canada, United States, Australia, New Zealand and Hong Kong remained under colonialism. On the other hand, Ethiopia, Skim, Tibet, Nepal, Bhutan, and Thailand were never colonies; however, these countries are the poorest in the world.
In fact, virtually all- African countries are known to be in poverty and their people experienced very poor living conditions, but the vast majority of the people wallow in abject poverty. For instance, Sudan according to the NBHS, 46.5 percent of households lives below the poverty line. Is the poverty prevailing? Are poor themselves responsible for their plight? Have they been made a poor decision? Are their governments accountable for their predicament? These socioeconomic factors of poverty also harm the development of the country but on the other hand, there are also such aspects that cause poverty which needs to be discussed in details. Therefore, along with economic perspectives, social and political factors may be and often are equally responsible for poverty.

Generally, in Africa macroeconomic indicators are so unfavorable which hinder the development of the country as well as poverty reduction. For instance, Sudan official statistics show that economic growth has not always been accompanied by a decrease in unemployment and poverty (Osinubi, 2005). Evidence from available surveys on changes in asset ownership and social indicators over the 1990s suggests that economic growth has not been distributed equitably. Non-monetary indicators show inequalities related to respect to gender, rural-urban residence, regional or state location, and IDPs versus settled residents. IDPs are more impoverished than people living in residential communities, have fewer assets, and are more vulnerable to famines (Central Bureau Statistics, 2003). Sudan is a heavily indebted poor country. Therefore, Sudan’s access to external financing is limited, and it is vital for Sudan’s further social development plan to come to terms with the debt situation.

2.2 Conceptual of Poverty

The issue of poverty has been for a long time at the heart of development efforts among the nations of the world; both poor and rich alike. It is a cardinal focus of the millennium development goal; the first of the eight goals which are to half the proportion of the world's population living below US$1 per day by the year, 2015 (Sabir, 2012). Poverty is a multidimensional concept and has been viewed in different ways by many scholars. However, what appears to be a common view is that poverty exists when one is unable to satisfy some certain basic requirements. Of course, the subject of "basic needs or requirements" is also debatable. However, irrespective of how poverty is conceived, poverty analysis begins with the recognition of basic needs, what constitutes these basic requirements as well as the notion of deprivation. While it is being regarded as the outcome of the interrelationship between the socio-economic, political and actions that eventually result in human deprivation and deterioration in the living conditions of the people (Sackey, 2005); it is regarded as lack of access to basic necessities of life such as food, water, shelter, health, and clothing as well as inaccessibility to the social, cultural and means necessary to guarantee, productivity, social reproduction, and everyday life of the society (Williamson, 2000). In attempting to summarise the definition of poverty, Englama and Bamidele (1997) asserted that poverty in both relative and absolute terms refers to a circumstance where a person is not able to fend or provide sufficiently for his or her necessities or fundamental human requirements such as clothing and decent accommodation, food, the fulfillment of social and economic responsibilities, non-access to productive employment, lack of skills, resources and confidence; and has restricted admission to economic and social infrastructure. These include access to health, education, potable water, sanitation, and roads. These preclude the person from advancing in welfare which is limited by the scarce availability of economic and social infrastructure. They concluded by terming this situation as being subject to a "lack of capabilities" (Englama & Bamidele, 1997).

The definition of poverty may vary from country to country. There are two kinds of poverty absolute or relative.

In Sudan, both absolute and relative poverty exists. Absolute poverty refers to the lack of basic needs, education health, clothing shelter etc. Relative poverty refers to the lack of a socially acceptable level of income or other resources as compared to other countries or societies. Absolute poverty can be alleviated but relative poverty is a vital concept and exists in all parts of the world. It involves a comparison between groups. The poverty largely measures in monetary terms (Sabir, 2012).

Poor people can be classified as groups of people who live in poverty. They can also be described as the disadvantaged groups who are marked by deprivation particularly of income and other basic needs of life. Poor people are those whose standard of living as measured by income or consumption is lower than the poverty line (Achial et al., 2011). Poverty, as evaluated by income or expenditure, is mainly used to identify households that are poor and this approach has become the preferred indicator for the quantitative measurement of poverty and living standards all over the world (Akerele, Momoh, & Ashaolu, 2012).

According to (Foster, Greer & Thorbeck., 1984) poverty index, among others, have been widely doped in the measurement of income poverty by many development scholars across over the world. It has contributed immensely towards the design, implementation and appraisal of prominent development programmers (Foster et al., 2010) in both developed and developing countries. Consequently, this study relies on the Foster, Greer, Thorbecke (FGT) poverty index because of its theoretical soundness, ease of application and its relevance within the domain of poverty and development policies.

Generally, there are two economic indicators are used in third world countries. One is poverty line approach, in terms of monetary indicator such as income and consumption. In this regard, Head Count Ratio (HCR) is used to define poverty on a national level such as US $1 or the US $2 per head per day. People who are living below the poverty line of US$1 or US$2 per head per day are considered poor. The other method is used to measure poverty is non-monetary which is called Unsatisfied Basic Human Needs (UBN). In this aspect, people are not able to have access to basic needs such as housing, basic health services and education (O’Hare et al. 2007).

Poverty has its spread all over the developed and developing countries and is found both in the rural and urban areas. The incidence of poverty is much higher in the rural areas but its impact in the urban areas is nonetheless considerable. The incidence of poverty (also referred to as headcount ratio) can be defined as the proportion/percentage of the population who are living in poverty (Foster et al., 2010). It is admeasuring that describes the rate or prevalence of poverty in certain population groups. Other poverty measures include poverty gap and severity of poverty. The poverty gap presumes that not all people are alike. The index basically differentiates among the poor. It provides information on how far an average poor is away from the poverty line. The emphasis of the Severity measure is on the status of the poorest of the poor (Foster et al., 2010).

2.2.1 Poverty line

Poverty lines are the starting point of every point of analysis. They are usually based on income and consumption data. According to the (Coudouel & Wodon, 2002), once an aggregate income, consumption, or non-monetary measure is defined at the household or individual level, the next step is to define one or more poverty lines. Poverty lines are cutoff points separating the poor from the non-poor. They can be monetary (for example, a certain level of consumption) or nonmonetary (for instance, a certain level of literacy). The use of multiple lines can help in distinguishing among different levels of poverty. There are two main ways of setting poverty lines-relative and absolute:

Relative poverty lines: These are defined in relation to the overall distribution of income or consumption in a country; for example, the poverty line could be set at 50 % of the country's mean income or consumption. Also, another visible, relative poverty refers to the position of an individual or household compared with the average income in the country, such as the poverty line set at the 40th percentile of the distribution. Relative poverty varies with the level of the average income (Busisa, 2011).

Absolute poverty lines: These are anchored in some absolute standard of what households should be able to count on in order to meet their basic needs. For monetary measures, these absolute poverty lines are often based on estimates of the cost of basic food needs, that is, the cost of a nutritional basket considered minimal for the health of a typical family, to which a provision is added for nonfood needs. Form another perspective, absolute poverty is the minimum basket of resources in which one needs to survive. In Sudan, the person whose income cannot meet daily intake 2,110 calories per day per person is considered below the poverty line. The relative poverty is living conditions and resources in the society in relations to others. Thus absolute poverty is hunger, deprivation, and lack of education, ill health, and suffering. On the other hand, relative poverty is the unequal distribution of resources and associated with a matter of social equity. It related to the average income of the society and social exclusion (Mubasher, 2009).

Considering that large parts of the populations of developing countries survive with the bare minimum or less, reliance on an absolute rather than a relative poverty line often proves to be more relevant.

Alternative poverty: lines are also sometimes used. They can be set on the basis of subjective or self reported measures of poverty. Moreover, absolute and relative poverty lines can be combined. This technique allows for taking into account inequality and the relative position of households while recognizing the importance of an absolute minimum below which livelihood is not possible. When deciding on the weight to give to the two lines when combining them, one can use information contained in the consumption or income data and information from qualitative data (if the qualitative data show that people consider a specific good to be a basic need, the elasticity of ownership of that good to income can be used).

The choice of a poverty line is ultimately arbitrary. In order to ensure wide understanding and wide acceptance of a poverty line, it is important that the poverty line chosen resonate with social norms, with the common understanding of what represents a minimum. For example, in some countries, it might make sense to use the minimum wage or the value of some existing benefit that is widely known and recognized as representing a minimum. Using qualitative data could also prove beneficial in deciding what goods would go in the basket of basic needs for use in constructing an absolute poverty line.

2.2.2 Poverty Measures

The poverty measure itself is a statistical function that translates the comparison of the indicator of household well-being and the chosen poverty line into one aggregate number for the population as a whole or a population subgroup. Many alternative measures exist, but the three measures described are most commonly used according to (Coudouel, Hentschel, & Wodon, 2002).

Their studies used the P-alpha equation of Foster-Greer and Thorbecke (FGT) to assess poverty incidence, gap, and severity. Both sectoral and subsector poverty indicators were calculated using the following P-alpha equation

$$PG_a = \frac{1}{N}\sum_{i=1}^{q}\left[\frac{z-y_i}{z}\right]^{\alpha} \dots\dots\dots\dots\dots (2.1)$$

Where: N sample size, Z poverty line, Y average income and α poverty aversion which has the value of 0, 1 and 2.

The incidence of poverty (headcount index): This is the share of the population whose income or consumption is below the poverty line, that is, the share of the population that cannot afford to buy a basic basket of goods. An analyst using several poverty lines say, one for poverty and one for extreme poverty, can estimate the incidence of both poverty and extreme poverty. Similarly, for nonmonetary indicators, the incidence of poverty measures the share of the population that does not reach the defined threshold (for instance, the percentage of the population with less than three years of education).

The study used this measure to capture the extent of poverty.

$$HCI_0 = {}^{q}/_{N} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2.2)$$

The depth of poverty (poverty gap): This provides information regarding how far off households are from the poverty line. This measure captures the mean aggregate income or consumption shortfall relative to the poverty line across the whole population. It is obtained by adding up all the shortfalls of the poor (assuming that the non poor have a shortfall of zero) and dividing the total by the population. In other words, it estimates the total resources needed to bring all the poor to the level of the poverty line (divided by the number of individuals in the population).

This measure can also be used for nonmonetary indicators, provided that the measure of the distance is meaningful. The poverty gap in education could be the number of years of education needed or required to reach a defined threshold. In some cases, though, the measure does not make sense or is not quantifiable (for example, when indicators are binary, such as literacy, in which case only the concept of the headcount can be used).

The poverty gap can be used as a measure of the minimum amount of resources necessary to eradicate poverty, that is, the amount that one would have to transfer to the poor under perfect targeting (that is, each poor person getting exactly the amount he/she needs to be lifted out of poverty) to bring them all out of poverty.

$$PG_1 = \frac{1}{N}\sum_{i=1}^{q}\left[\frac{z-y_i}{z}\right]^1 \ldots\ldots\ldots\ldots\ldots (2.3)$$

Poverty severity (squared poverty gap): This takes into account not only the distance separating the poor from the poverty line (the poverty gap), but also the inequality among the poor. That is, a higher weight is placed on those households further away from the poverty line. As for the poverty gap measure, limitations apply to some of the nonmonetary indicators.

$$PG_2 = \frac{1}{N}\sum_{i=1}^{q}\left[\frac{z-y_i}{z}\right]^2 \ldots\ldots\ldots\ldots\ldots (2.4)$$

All of these measures can be calculated on a household basis, that is, by assessing the share of households that are below the poverty line in the case of the headcount index. However, it might be better to estimate the measures on a population basis in terms of individuals in order to take into account the number of individuals within each household.

The measures of depth and severity of poverty are important complements of the incidence of poverty. It might be the case that some groups have a high poverty incidence but low poverty gap (when numerous members are just below the poverty line), while other groups have a low poverty incidence but a high poverty gap for those who are poor (when relatively few members are below the poverty line but with extremely low levels of consumption or income).

According to the headcount, unskilled workers show the third highest poverty rate, while this group ranks fifth in poverty severity.

Comparing them with the herders shows that they have a higher risk of being in poverty but that their poverty tends to be less severe or deep. The types of interventions needed to help the two groups are therefore likely to be different.

Depth and severity might be particularly important for the evaluation of programs and policies. A program might be very effective at reducing the number of poor (the incidence of poverty) but might do so only by lifting those who were closest to the poverty line out of poverty (low impact on the poverty gap). Other interventions might better address the situation of the very poor but have a low impact on the overall incidence (if it brings the very poor closer to the poverty line but not above it).

2.3 Poverty in Sudan

Sudan shares its borders with nine countries. Population density varies widely across the country with 67% rural and 33% urban. More than 90% of the population suffers from poverty and food insecurity. With a total land area of 2.5 million km^2 Sudan is the largest country in Africa (Kong, & FAO, 2005). Sudan has had one of the highest growth rates amongst Sub-Saharan African countries and a rapidly rising per capita income, with per capita GDP of US$1,500. Nonetheless, the country's human development outcomes remain weak. Sudan ranks 154 out of 169 countries in UNDP's 2010 Human Development Index, especially relative to the fact that income per capita GDP exceeded $1,500 or roughly 25% higher than the Sub-Saharan Africa (SSA) average. In 2009, Sudan was the third largest producer of crude oil in SSA, behind Nigeria and Angola, although Sudan's production was only about 30% of Angola's.

Looking back 40 years or so, the living conditions of larger numbers of people were much better than today (El-Batthani et al. 1998. Abusin 2003). During this period, larger numbers of people were affected by poverty regardless of their mode of living and although data on poverty trends and magnitudes is lacking or at best scanty, preliminary studies show that the Sudanese economy has deteriorated steadily since the 1970s.

This is reflected in the decline in per capita income from $500 in the 1970s to $430 in 1980 to around $290 in 1998 (Abusin 2003& Eltigani, 1995).

A large number of poverty alleviation programmers have been implemented by successive Sudanese government since independence. But the majority of rural and estate people are yet under severe poverty-ridden conditions. A number of attempts were taken in the last two decades to quantify poverty in Sudan and to delineate its causes. These efforts were however handicapped by the limitations of data and the incomplete coverage of the survey on which they are based.

A serious limitation was the unavailability of a household budget survey on which to base the estimate of poverty and to gauge its changing trends since only one survey was carried out in 1967/68 and twenty years elapsed before the second one in 1978/80 and others in 1992 the ILO funded the migration and labor force survey. Also, in 1992 the Social Solidarity Fund funded the poverty line survey. The 2009 National Baseline Household Survey (NBHS), the first nationally representative household consumption survey conducted in Sudan since 1978, provides estimates for the various dimensions of poverty (CBS, 2010). Within the last three decades, estimates of national poverty rates show a steady increase in poverty from 50% in 1968 up to 75% in 1986 and as high as 80% in 1997. Based on recorded income, a 1992 study estimated poverty at about 86% (Awad, 1997).

Another study indicates that, between 1986 and 1996, the proportion of the population below the poverty line increased from 52.9% to 84.6% in urban areas and from 83.1% to 93.9% in rural areas (Kossaifi, 1998). It is worth noting that even in periods when the country was experiencing economic growth; poverty continued to escalate. For example, between 1991 and 1997 the economy had an average annual growth rate of 8% in real terms but poverty increased from 71% to 91% between 1990 and 1996 (Kossaifi, 1998). In recent years Sudan has experienced rapid economic growth but the effects of this growth on poverty have unfortunately not yet been documented.

The table (2.1) below shows that the trend of headcount index in Sudan has been increasing at an annual rate of 0.5%., the number of rural households have been growing at a rate equal to the rural population growth rate while the number of poor urban households have been growing at a rate higher than the urban population growth rate. The poverty gap ratio in the whole country has been decreasing at an annual growth rate of 0.64%. (Ali, 2002).

Table (2.1): Poverty in Sudan during 1968 – 1993

Indicator	1968	1978	Growth Rate	1986	1993	Growth Rate
HCI						
Rural	62.68	64.17	0.23	83.12	93.16	1.26
Urban	15.9	20.51	2.58	52.86	84.43	3.9
Sudan	51.59	54.26	0.5	77.8	91.41	1.7
PGI						
Rural	28.11	30.56	0.84	51.67	62.61	1.4
Urban	4.56	8.58	6.53	24.38	47.78	2.9
Sudan	24.66	23.12	-0.64	45.43	59.35	1.7

Source: *Estimations made by Ali Abdel Gadir "SAPs and Poverty in Sudan 2002"*

Poverty assessment in Sudan has been limited, but studies provide evidence of high-income poverty. Earlier analyses by region revealed very high, rising poverty incidence between 1990 and 1996; ranging in its urban dimension from 87-91% to 77-93% and in its rural dimension from 55-77% to 80-97% from 1990 to 1996 in the six regions of North Sudan. Rural areas had witnessed escalating poverty incidence during the first half of the 1990s rendering poverty as a dominantly rural phenomenon (Hamid, 2010) see the following table (2.2).

Table (2.2) Poverty Incidence by region 1990-1996 (%)

Region	1990		1993		1996	
	Urban	Rural	Urban	Rural	Urban	Rural
Darfur	97	55	89	89	89	97
Kordofan	91	77	91	84	87	96
Central	88	67	89	83	93	91
Eastern	89	60	82	81	88	94
Northern	89	56	91	80	90	93
Khartoum	84	56	75	64	77	80

Source: Hamid Faki, Eltahir M. Nur and AbdelazizHashim (ICARDA forthcoming).

The bulk of Sudanese was in absolute poverty ranging between 85% and 94%, and the rich-poor gap is growing. 86.5% of urban dwellers are classified as “poor”.

The growth rate of the number of poor families was estimated to be more than 4percent annually (Ibrahim, 2003).

According to more recent (2009) poverty analysis with wide coverage by state and based on consumption aggregates of five main components (food, non-food, durable goods, housing, and energy) puts North Sudan at an overall poverty level of 46.5% of households in Sudan live below the poverty line. This represents approximately 14.4 million people. The poverty line is defined as persons with the value of monthly total consumption below SDG 114 (calculated using 2400 calories per person per day as the daily energy intake threshold).

Poverty is still widespread in Sudan in spite of the efforts exerted to reduce its incidence. According to (CBS, 2009), almost half of the population of Northern Sudan is found to fall below the poverty line, with 26.5% of the urban population and 57.6% of the rural population. Out of the population of Southern Sudan, 50.6% is found to fall below the poverty line, with 24.4% of the urban population and 55.4% of the rural population. Khartoum is the region with the lowest poverty incidence, followed by Northern. Eastern and Central rank third, while Kordofan and Darfur are the poorest regions. Poverty levels vary greatly by state. The incidence of poverty ranges from one fourth in Khartoum to more than two thirds in Northern Darfur (CBS, 2010). However, the rural incidence of poverty (83.1) remained higher than the incidence of urban poverty (53%). Poverty rates are substantially higher among the e rural dweller, with the percent of households below the poverty line compared to percent of the urban population. In Northern Sudan with 58% of the population living in rural areas (excluding the nomads), the share of the Rural poor in total poor is about 75%. With rural areas predominantly agricultural, this poverty is largest among agricultural households, show the following table (2.3).

Table (2.3): Poverty profile 2009

	Incidence	**Poverty** Poverty gap	Severity	**Poverty gap** among the poor	**Population** (%)	**Poor** (%)
Northern Sudan	46.5	16.2	7.8	34.8	100	100
Urban	26.5	7.1	2.7	26.6	35.6	20.3
Rural	57.6	21.3	10.6	36.9	64.4	79.7
Northern	33.7	9.4	3.8	28	6.4	4.7
Eastern	46.3	17.7	9	38.2	14.3	14.2
Khartoum	26	6.4	2.4	24.7	18.7	10.4
Central	45.4	13.8	6.1	30.4	26.2	25.5
Kordofan	58.7	23.1	11.7	39.3	14.3	18.1
Darfur	62.7	24.6	12.6	39.3	20.1	27.1
Northern	36.2	10.5	4.2	29.1	2.4	1.9
River Nile	32.2	8.8	3.5	27.3	4	2.8
Red Sea	57.7	24.9	13.7	43.1	3.6	4.4
Kassala	36.3	14.7	8	40.6	5.9	4.6
Al-Gadarif	50.1	15.9	6.7	31.8	4.8	5.2
Khartoum	26	6.4	2.4	24.7	18.7	10.4
Al-Gezira	37.8	10.1	4.1	26.6	12.2	9.9
White Nile	55.5	17.6	7.8	31.7	6.4	7.6
Sinnar	44.1	14	6.4	31.7	4.5	4.3
Blue Nile	56.5	20.6	9.9	36.5	3.1	3.7
Northern Kordofan	57.9	24.6	13.1	42.5	8.9	11
Southern Kordofan	60	20.7	9.4	34.5	5.5	7.1
Northern Darfur	69.4	27.4	14.2	39.6	5.9	8.7
Western Darfur	55.6	19.8	8.9	35.6	3.2	3.8
Southern Darfur	61.2	24.5	12.7	40.1	11.1	14.6

Source: CBS, NBHS 2009

Poverty assessment focuses not only on the lack of material deprivation but also on the deprivation of non-tangible services such as education, health and shelter (Semasinghe, 2009).

The economic status of people shapes their opportunities, decisions, and expectations in different life spheres and stages. All that impacts people out of poverty is a consequence of economic and demographic conditions, as well as social, political and historical reasons. The degree of poverty that a society might experience depends on the volume and distribution of resources and on the size and distribution of the population among households. These two basic determinants of poverty, however, are not independent. On the one hand, the size and age structure of a population are consequences of fertility decisions taken over past decades which were influenced by the prevailing economic conditions. On the other hand, the volume of resources available today is influenced by the size and age composition and productivity of the labor force.

It is most important to note that the poverty trends differ very slightly and sometimes vary greatly between groups. In general terms, the number of the poor people in rural areas has increased with a rate nearly equal to the rate of population increase. And the number of the poor urban household has increased at a higher rate than the urban population growth rate (Obwona, & Guloba, 2009).

Poverty Profile 2015: The total estimated population covered by the NHBPS 2014/15 is 34.2 million persons distributed on 6 million households. The survey does not include population groups such as nomads, people living in camps and homeless people etc and can thus not directly be compared to the population counted in the Census of 2008 or updates hereof. The results of the NHBPS survey 2015 show that the annual per capita consumption in Sudan was Sudanese Pounds (SDG) 6,082. Urban areas displayed average consumption levels higher than rural areas, at SDG 7,149 and SDG 5,509 respectively. Among States, average consumption was the highest in Khartoum, followed by Northern and River Nile. States of Darfur and Kordofan regions showed the lowest level.

The prevalence of the global poverty in Sudan was 36.1% of the sample households live below the poverty line with an average poverty gap of 10.3. One in four Sudanese falls below the extreme poverty line (25%).

The poverty line computed from Sudan included National Household Budget and Poverty Survey data using the cost of basic needs method. The decline in the poverty line in Sudan 2015 is due to the change in calorie measurement method. In the past, international standards were used for calorie measurement, but in this study, local calories 2110 were used to measure the poverty line, according to a briefing of the Central Bureau of Statistics. While, find that the highest poverty in Central Darfur and South Kordofan at 67.2% and 67% respectively, but the lowest in the northern state at 12.2%. To struggle poverty in Sudan, the World Bank agreed to provide $100 million in order to establish development projects in Sudan until 2019. Sudan's state minister predicted that the economy of Sudan would grow by only 0.2% per year2. However, the situation remained unchanged in the economy, but now a day's worse than it was.

2.3.1 Underlying Causes for Poverty in Sudan

Poverty is a multidimensional and complex phenomenon and is related not only to the income or consume, considered as monetary dimension of poverty, but also to non-monetary dimensions such as education, health, gender equality, water supply, etc. Poverty is caused by many factors and brings several effects which influence the lives of people considered to be poor. The influence of the factors varies from one place to another, because many countries have different development possibilities. The influential factors of poverty level are not only economical, but also social, political, cultural, geographical, etc.

According to the (IMF, 2013) Poverty is caused by many factors and brings several effects which influence the lives of people considered to be poor. There are many roots that account for persistent poverty in Sudan but the main roots include: first; the long drawn out civil conflicts in southern, western and eastern Sudan that diverted attention and resources from development to fighting wars, impaired social capital, and good governance and destroyed human and physical capital. The lack of durable peace and security dissuaded households and firms from making investments in human and physical capital for the future Scand; The urban bias of development policies and programs in the past that neglected efforts to broadly increase the productivity of rural factors of production, particularly in the sphere of rain-fed agriculture; Third; lack of a coherent poverty reduction effort and a sustained reform to promote shared growth and diversify the economy, hence the rising unemployment; Low allocation of public resources to poverty reduction priorities, particularly agricultural development and human development, and the absence of development partners to compensate for the under-spending; Fours; the concentration of socio-economic development in a few areas; and.

And the burden of an unsustainable external debt, the long economic international sanctions and isolation that held up access to the international debt relief initiatives such as HIPC and to concessional financial assistance.

Poverty may be caused or exacerbated by:

- The lack of capacity of the poor to influence social processes, public policy choices and resource allocations.
- Low capacities through lack of education, vocational skills, entrepreneurial abilities, poor health and poor quality of life.
- The disadvantaged position of women in society.
- Exposure to risks through lack of financial, social or physical security.
- Low levels of consumption through lack of access to capital, social assets, and land and market opportunities.
- Exposure to shocks due to limited use of technology to stem effects of drought, floods, armyworms, crop pests, crop diseases, and environmental degradation.
- Inadequate environmental protection measures.
- Lack of macroeconomic stability that erodes the resources of the poor through inflation and other variables.
- The inability of the national economy to optimize benefits within the global system.
- Habits and conventions based upon superstition and myths giving rise to anti-social behavior.

- Other factors leading to vulnerability and exclusion.

Appending of that, Poverty in Sudan is caused by corruption and poor governance, poor land utilization and land tenure system, civil wars and unending political conflicts, poor infrastructure, diseases and poor health facilities, the World Bank and IMF policies, among others. According to the World Bank (1990), and the United Nations (1995), poverty has various manifestations which include the lack of income and productive resources sufficient to ensure sustainable livelihood, hunger, and malnutrition, ill health, limited or lack of access to education and other basic services, increased morbidity and mortality from illness, homelessness, inadequate, unsafe and degraded environment, social discrimination and exclusion. It is also characterized by lack of participation in decision making in civil, social and cultural life (World Bank, 2001). In their discussions of the factors that cause poverty, De Haan (2000) and Sindzingre (2000) noted that poverty could also be caused by general exclusion of the people from social life. To them exclusion reflects discrimination, which is a process that denies individuals from full participation in material exchange or interaction. The concept is tied to exclusion from the labour market, long-term unemployment and the destruction of the social links and integration that usually accompany work.

2.3.2 The Choice of the Monetary Indicator

The main decision in poverty estimation is to choose between income and consumption as the welfare indicator to determine poverty. Consumption is the preferred measure because it is likely to be a more useful and accurate measure of living standards than income. This preference of consumption over income is based on both theoretical and practical issues (Deaton & Zaidi (2002) and Hentschel & Lanjouw (1996)).

The first theoretical consideration is that both consumption and income can be approximations to utility, even though they are different concepts (Magrabi, 1991). Consumption measures what individuals have actually acquired, while income, together with assets, measures the potential claims of a person. Secondly, the time period over which living standards are to be measured is important: if one is using a long-term perspective as in a lifetime period, both should be the same and the choice does not matter.

In the short-run though, say a year, consumption is likely to be more stable than income. Households are often able to smooth out their consumption, which may reflect access to credit or savings as well as information on future streams of income. Consumption is also less affected by seasonal patterns than income: for example, in agricultural economies, income is more volatile and affected by growing and harvest seasons, hence relying on that indicator might under or overestimate significantly living standards.

There are also practical arguments to take into account. First, consumption is generally an easier concept than income for the respondents to grasp, especially if the latter is from self-employment or family-owned businesses. For instance, workers in formal sectors of the economy will have no problem in reporting accurately their main source of income, i.e., their wage or salary. But self-employed persons in informal sectors, or engaged in agriculture, will have a harder time coming up with a precise measure of their income. Often in these cases, household and business transactions are intertwined. Besides, as was mentioned before, seasonal considerations are to be included to estimate an annual income figure. Finally, we also need to consider the degree of reliability of the information. Households are less reluctant to share information on consumption than on income. They may be afraid than income information will be used for different purposes, say taxes, or they may just consider income questions as too intrusive. It is also likely that household members know more about the household consumption than the level and sources of household income.

2.3.3 The Construction of the Consumption Aggregate

Creating the consumption aggregate is also guided by theoretical and practical considerations. In the case of the NBHS, the focus will be on the consumption aggregate of the household in the last year. First, it must be as comprehensive as possible given the available information. Omitting some components assumes that they do not contribute to people's welfare or that they do not affect the rankings of individuals.

Second, market and non-market transactions are to be included, which means that purchases are not the sole component of the indicator. Third, expenditure is not consumption. For perishable goods, mostly food, it is usual to assume that all purchases are consumed.

But for other goods and services, such as housing or durable goods, corrections have to be made. Lastly, the consumption aggregate comprises five main components: food, non-food, durable goods, housing and energy.

2.4 The Poverty Line in Sudan

According to (Ravallion, (1998) & Ravallion, (1996)) the poverty line can be defined as the monetary cost to a given person, at a given place and time, of a reference level of welfare. If a person does not attain that minimum level of standard of living, she will be considered poor. Implementing this definition is, however, not straight-forward because considerable disagreement could be encountered at determining both the minimum level of welfare and the estimated cost of achieving that level. In addition, setting poverty lines could be a very controversial issue because of its potential effects on monitoring poverty and policy-making decisions.

It will be assumed that the level of welfare implied by the poverty line should enable the individual to achieve certain capabilities, which include a healthy and active life and a full participation in society. The poverty line will be absolute because it fixes this given welfare level, or standard of living, over the domain of analysis. This guarantees that comparisons across individuals will be consistent, for instance, two persons with the same welfare level will be treated the same way regardless of the location where they live. Second, the reference utility level has been anchored to certain attainments, in this particular case to the attainment of the necessary calories to have a healthy and active life. Finally, the poverty line will be set as the minimum cost of achieving that requirement.

The Cost of Basic Needs method was employed to estimate the nutrition-based poverty line. This approach calculates the cost of obtaining a consumption bundle believed to be adequate for basic consumption needs. If a person cannot afford the cost of the basket, this person will be considered to be poor.

First, it shall be kept in mind that the poverty status focuses on whether the person has the means to acquire the consumption bundle and not on whether its actual consumption met those requirements. Second, nutritional references are used to set the utility level but nutritional status is not the welfare indicator. Otherwise, it will suffice to calculate caloric intakes and compare them against the nutritional threshold. Third, the consumption basket can be set normatively or to reflect prevailing consumption patterns. The latter is undoubtedly a better alternative. Lastly, the poverty line comprises two main components: food and non-food. According to the household survey, the monthly per capita consumption in Northern Sudan in 2009 was SDG 148. Urban areas display consumption levels significantly higher than in rural areas (SDG 197 and 122 respectively).

Food poverty line in 2009 was estimated at SDG 69 per person per month and estimated at SDG 45 per person per month for non food and total consumption 114 SDG per person per month in Sudan, see following table (2.4).

Table (2.4): Poverty Line per Person per Month 2009

	SDG	%
Food	69	61
Non-food	45	39
Total	114	100

Source: CBS, NBHS 2009

A specific poverty line is estimated for each of rural and urban area to take into account Sudanese cost of living in various area of residence. Food poverty line in 2015 was estimated at SDG 2,966 in urban areas and SDG 2,698 in rural areas. The global poverty line was estimated at SDG 5,110 per person annually in urban area and SDG 4,044 in rural area in the following table (2.5).

Table (2.5): Poverty Lines per Person per Year in 2015

Area	Food poverty line	Extreme poverty line	Global poverty line
Urban	2966	4124	5110
Rural	2698	3605	4044

Source: CBS, NBHS 2015

The percentage of the population with a consumption level below the poverty line referred to "incidence of poverty". The prevalence of the global poverty in Sudan was 36.1%. One in four Sudanese falls below the extreme poverty line (25%).

Table (2.6) Poverty incidence in 2015

Area	Population below the extreme poverty line	Population below the global poverty line
Urban	22.6%	37.3%
Rural	26.5%	35.5%
Sudan	25.2%	36.1%

Source: CBS, NBHPS 2015

2.5 Poverty Measures in Sudan

The literature on poverty measurement is extensive, but attention will focus on the class of poverty measures proposed by Foster, Greer, and Thorbecke (1984). This family of measures can be summarized by the following equation:

$$P_\alpha = (1/n)\sum_{i=1}^{q}\left(\frac{z-y_i}{z}\right)^{\alpha} \qquad \ldots\ldots\ \ \ldots\ldots \quad (2.5)$$

Where α is some non-negative parameter, *z* is the poverty line, *y* denotes consumption, *i* represents individuals, *n* is the total number of individuals in the population, and *q* is the number of individuals with consumption below the poverty line.

The headcount index (α=0) gives the share of the poor in the total population, that is, it measures the percentage of the population whose consumption is below the poverty line. This is the most widely used poverty measure mainly because it is very simple to understand and easy to interpret. However, it has some limitations. It takes into account neither how close nor far the consumption levels of the poor are with respect to the poverty line, nor the distribution of consumption among the poor. The poverty gap ($\alpha=1$) is the average consumption shortfall of the population relative to the poverty line. Since the greater the shortfall, the higher the gap, this measure overcomes the first limitation of the headcount.

Finally, the severity of poverty ($\alpha=2$) is sensitive to the distribution of consumption among the poor, a transfer from a poor person to somebody less poor may leave unaffected the headcount or the poverty gap but will increase this measure. The larger the poverty gap is, the higher the weight it carries.

These measures satisfy some convenient properties. First, they are able to combine individual indicators of welfare into aggregate measures of poverty. Second, they are additive in the sense that the aggregate poverty level is equal to the population-weighted sum of the poverty levels of all subgroups of the population. Third, the poverty gap and the severity of poverty satisfy the monotonicity axiom, which states that even if the number of the poor is the same, but there is a welfare reduction in a poor household, the measure of poverty should increase. And fourth, the severity of poverty will also comply with the transfer axiom: it is not only the average welfare of the poor that influences the level of poverty, but also its distribution. In particular, if there is a transfer from one poor household to a richer household, the degree of poverty should increase (Sen., 1976).

2.6 Concept of Socio-economic Factors

Poverty is generally examined in academic literature from two major angles: Poverty as determined by micro-level household and individual characteristics such as a household size, education of the head of a household, etc. on the one hand, and aggregate macro-level economic indicators measured on a country level on the other. Micro-level studies typically make use of household and individual surveys, while macro-level analysis usually employs country-specific economic and social indicators. Even though the two approaches are rather different, they sometimes measure the same factors and may even be combined in one comprehensive study.

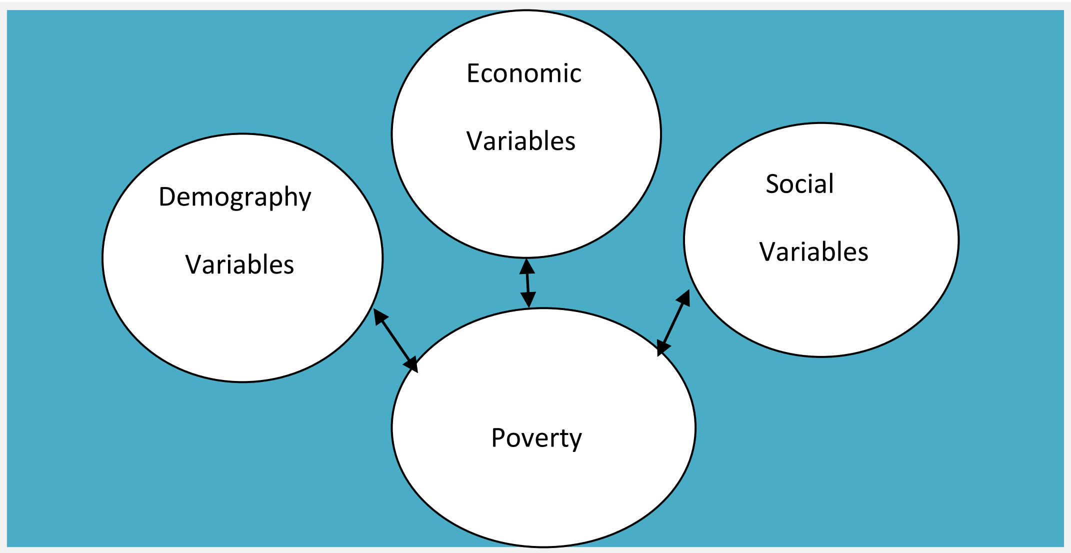
The studies on the micro-level are especially relevant for the purposes of this research so I will focus this overview on this type of approach in the literature, as this study has an objective to impact of socioeconomic factors on poverty. Micro regressions are useful for describing the distinctive features of multidimensional poverty profiles across households (in a given country) or to understand the determinants of poverty, (Alkire, & Foster et al., 2015).

Some important socioeconomic variables would include such household migration, demographic characteristics of household members, education, labor force, housing characteristics, livelihood and assets, transfers to the household, consumption, agriculture, and income.

2.6.1 Interrelationship between Poverty and Socioeconomic Variables

There are some socioeconomic and demographic factors in micro or macro- level, which play a significant role in determining the poverty status. Our study is concerned with microeconomic variables and characteristics. It would be appropriate to provide a brief explanation of how these factors are correlated with poverty.

Figure (2.1) Relationship between Poverty and Demography, Social and Economic



Source: Researcher own Construction (2017)

2.6.4 Demographic Characteristics of Households

Aside from economic and social indicators, we make use of significant demographic indicators to characterize poverty and household living standards. The demographic characteristics of the household can be broadly classified into, as follows:

i. Household Size and Structure

This indicator is an important one as it shows a possible correlation between the level of poverty and household composition. Household composition, in terms of the size of the household and characteristics of its members (such as age), is often quite different for poor and non poor households. The Sudan Integrated Censuses of 2008 shows that the average family size of 6 persons. Generally, it is recognized that more healthy, educated, and adult members in a household contribute to their income levels and reduce poverty; if household members are not adult and educated, they can become the cause of poverty. It is hypothesized that the larger the household size, the higher the level of poverty incidence, and vice versa.

ii. Dependency Ratio

For a given household size, a larger number of children and elderly members would imply a smaller number of earners in the household. In the present analysis, the dependency ratio is calculated as the ratio of the number of members below 15 and over 64 to other household members. Furthermore, child and older member dependency ratios are also calculated using the same formula. This ratio allows us to measure the burden on members of the labor force within the household. One might expect that a high dependency ratio would be correlated positively with the level of rural household poverty.

iii. Female-Male Ratio

The female-male ratio or sex ratio is important in a household in determining the attitude toward work. Although not to be assumed a generalization, female household members in Sudan are often constrained by cultural norms from working outside their household. This suggests that a high female-male ratio might be related to household poverty.

iv. Age and Gender of Household Head

The age and gender of the household head are also important in determining the attitude toward employment. It is widely believed that the age and gender of the household head significantly influence poverty. The age of the household head has a similar role to sex composition, as discussed above.

2.6.2 Social Determinants

Aside from economic variables, there are some social indicators are correlated with poverty and household living standards. The social indicators generally selected are household characteristics of education, marital status, and work. In the following sub-sections how these indicators relate to poverty will be discussed.

i. Education

According to human capital models, education is an important dimension of the non-homogeneity of labor. High educational attainment may imply a greater set of employment opportunities and specifically in the rural context, a better awareness of the full potential of new agricultural technologies and associated agricultural practices. Four types of indicators are normally used to characterize living standards. These include the number of household members, level of education (literacy rate, with poor households having lower literacy), availability of educational services (primary and secondary schools), the use Socio-Economic Determinants and Household Poverty Status of these services by members of poor and nonpoor households (children's enrollment in school, dropout rate of children by age and gender and reasons for dropping out, percentage of children who are older than the normal age for their level of education and average spending on education per child registered) and educational codes.

The educational index is constructed by dividing the total number of educational points by household size. This variable is considered a major cause of poverty and points are given to those household members who have completed their education up to secondary level or higher; these members are observed as being older than 14 years and are assumed to be adults. In view of its potential role, we hypothesize a positive relationship with per capita income, and a negative one with poverty incidence (Chaudhry et al., 2009).

ii. Marital Status

It has been posited that marriage brings an array of benefits (Waite and Gallagher, 2000): in economic terms, since marriage generally adds a potential earner to the household, it seems obvious that marriage should increase the economic well-being of members of the family, including the children. Married women living in male-headed households have the prospect of enjoying larger family income because these families have a larger number of earning members and especially a larger number of earning male members. A long-term marital relationship may also mean higher permanent income and a larger buildup of consumer durables, factors that could limit the extent of economic hardship experienced in downturns in the economy. In addition, married couples may be more easily able to draw on relatives for help in difficult situations (Lerman, 2002).

2.6.3 Economic Determinants

Economic determinants include labor force, housing characteristics, livelihood and assets, transfers to the household, consumption, agriculture, and income. The poverty of Sudan is highly characterized by unemployment even though there are other causing factors. Generally, one can explicitly understand that conceptual linkage among the economic growth employment, population, and poverty.

Average dependency ratio due to the high population is high in developing countries in which labor productivity would be low because of inadequate nutritional food, health, and education. General theory tells us that lower the labor productivity lower the economic growth and higher the unemployment and poverty.

i. Household Employment

There are several indicators that determine household employment. Within this array of indicators, economists focus on the rate of participation in the labor participation rate is the first of the two employment variables used in the analysis. According to Lipton (1983), the higher the illness, disability, income per capita and intensity in customs, status, the general welfare level and asset holdings, the lower the participation rate in LDCs. In comparing the non poor and poor, the positive incentive given by poverty to participation outweighs the negative effect on it; hence, the poor participate more than the non poor, (Chaudhry, 2009).

ii. Household Incomes

Income represents a very important area of consideration when characterizing the poor. The level of income is important not only for the households, but its distribution among household members and various socioeconomic groups. Income is difficult to define as it includes several components of which only some are monetary (for example, farm households consume most of their production onsite). Additionally, individuals tend to make false declarations about their income level, which is generally underestimated. It is possible in part to correct these declarations but only at the cost of carrying out a large-scale data-gathering operation on economic activities, the cost of production, factor inputs, and the prices of products. Given these limitations and the fact that savings are low, even zero, there is often a tendency to use a household's total spending as an approximation of its disposable income. Here, we calculate per capita expenditure per year as a proxy for household income. The household classified as either poor or non-poor based on their per capita expenditure.

iii. Household Property and Assets

The property of a household includes its tangible goods (land, cultivated areas, livestock population, agricultural equipment, machinery, buildings, household appliances, and other durable goods) and its financial assets (liquid assets and other financial assets). These indicators are of interest as they represent the household's inventory of wealth and therefore affect its income flow. Furthermore, certain households, especially in the rural areas of Sudan, might be poor in terms of income but wealthy when their property is taken into consideration. This class of poverty is called secondary poverty by Rowntree (1901), as it applies to those who appear to have resources but have not been able to utilize them to raise themselves above the subsistence level. However, we and assets under the following head.

• Landholdings

The ownership of agricultural land is considered the main factor that can extricate a household/individual from poverty. The variable or characteristic used in this study is the extent of landholdings per household in acres. This incorporates owner-cum-sharecroppers as well as sharecroppers. On the basis of the role it plays in a rural economy, we hypothesize a positive relation to the per capita income variable. Some technological and agricultural input variables (use of tractor, HYVs, fertilizer, and pesticides, and irrigation water, etc.) are also associated with landholdings and have also a positive relation to per capita income.

• Livestock Population

The livestock sector is an important sector of the rural economy in Sudan. The contribution of the livestock sector toward family income is quite substantial. In the present study, this form of property or asset is normally included and measured in monetary units. It also has a positive relationship with per capita income in our analysis.

• Physical Assets

Physical assets contribute significantly to per capita income. In the present study, physical assets occur in the form of agricultural equipment and machinery, i.e., tractors and accessories, etc., and household appliances such as electronic goods. These are measured in terms of the rupee value of total physical assets.

iv. Household Consumption and Spend

In this study, household spending on food and non-food items has been used as a dependent variable instead of income. This is because information and data on income are difficult to obtain especially in developing countries, and particularly among low-income groups who don't have sustained sources of income or can't recall correctly the amount of income.

In Sudan, a few studies have focused on factors affecting poverty at the district level in Sudan. They were not deepening the knowledge of the factors affecting the phenomenon of household poverty status. This rise to the need to find those variables that most causes poverty. The gab, most of the studies in Sudan depend on household income but in this study, spending on food and non-food items has been used instead of income. Therefore, information on income obtained via field surveys may give low-quality data which urges for the use of consumer spending as a better indicator for poverty measurement and for detecting causality relationships than income.

CHAPTER THREE

LOGISTIC REGRESSION

3.1 Preface

3.2 Generalized Linear Methods

3.3 Logistic Regression

3.4 Binary Logistic Regression Model

3.5 Model Selection Methods

3.6 Models Building Strategy

3.7 Estimation Binary Response Probabilities

3.8 Estimation the Model Parameters

3.9 Goodness of Fit Statistics

3.10 Model Validation

3.11 Residual Diagnostics

LOGISTIC REGRESSION

3.1 Preface

This chapter presents description of different logistic regression models techniques such as generalized linear model, multiple regression, probit model, binary logistic model, estimation and assumption and testing of the models.

3.2 Generalized Linear Models (GLM)

Generalized linear models are defined according to (Nelder, and Baker, 1972). The class of generalized linear models is an extension of traditional linear models that allows the mean of a population to depend on a linear predictor through a nonlinear link function and allows the response probability distribution to be any member of an exponential family of distributions. Many widely used statistical models are generalized linear models. These include classical linear models with normal errors, logistic and probit models for binary data, and log-linear models for multinomial data. Many other useful statistical models can be formulated as generalized linear models by the selection of an appropriate link function and response probability distribution. The linear regression model has found widespread application in the social sciences mainly due to its simple linear formulation, easy interpretation, and estimation. In monetary poverty analysis, linear regression analysis has been used to study the determinants of household consumption expenditures or to model the growth elasticity of per capita income or income poverty aggregates like the headcount ratio or the poverty gap index.

Hypothesis tests applied to the Generalized Linear Model do not require normality of the response variable, nor do they require homogeneity of variances. Hence, Generalized Linear Models can be used when response variables follow distributions other than the Normal distribution, and when variances are not constant. Parameter estimates are obtained using the principle of maximum likelihood; therefore hypothesis tests are based on comparisons of likelihoods or the deviances of nested models.

3.3 Logistic Regression

Logistic regression is used increasingly in a wide variety of applications. Early use was in biomedical studies but that past 20 years have also seen much use in social science research and marketing. Recently logistic regression has become a popular tool in business applications some credit – scoring applications use logistic regression to model the probability that a subject pays a bill on time may use predictors such as the size of the bill annual income acceptations and so on. The logistic regression is often preferred as a model for binary responses as it is appropriate for any kind of data: cross-sectional, prospective, and retrospective. The predictor variables can be numerical or categorical (including binary). Multinomial (aka polychotomous) logistic regression can be used when there are more than two possible outcomes for the response. But here the focus will be in the typical binary response version (Myers, Montgomery, Vining, and Robinson, 2012).

3.3.1Basic Concepts of Logistic Regression

The use of logistic regression model dates back to 1845. It first appeared during the mathematical studies for the population growth at that time (Cokluk, 2010). The term logistic regression analysis comes from logit transformation, which is applied to the dependent variable. This case, at the same time, causes certain differences both in estimation and interpretation (Hair, Black et al, 2006).

Logistic regression analysis is also called "Binary Logistic Regression Analysis", "Multinomial Logistic Regression Analysis" and "Ordinal Logistic Regression Analysis", depending on the scale type where the dependent variable is measured and the number of categories of the dependent variable. Logistic regression is divided into two: "univariate logistic regression" and "multivariate logistic regression" (Stephenson, 2008). Data related to confronted and researched cases in applied social sciences are mostly categorical (nominal) data with discrete value or data obtained by an ordinal scale. For instance, the household is poor or not poor (Cokluk, 2010).

The logistic regression model has become, in many fields, the standard method of data analysis concerned with describing the relationship between a response variable and one or more explanatory variables where the response variable follows a binomial distribution. Logistic regression sometimes called the logistic model or logit model analyzes the relationship between multiple independent variables and a categorical dependent variable and estimates the probability of occurrence of an event by fitting data to a logistic curve. There are two models of logistic regression, binary logistic regression, and multinomial logistic regression. Binary logistic regression is typically used when the dependent variable is dichotomous and the independent variables are either continuous or categorical. When the dependent variable is not dichotomous and is comprised of more than two categories, a multinomial logistic regression can be employed.

Logistical regression is regularly used rather than discriminant analysis when there are only two categories of the dependent variable. Logistic regression is also easier to use with SPSS than discriminant analysis when there is a mixture of numerical and categorical independent variable's because it includes procedures for generating the necessary dummy variables automatically, requires fewer assumptions, and is more statistically robust. Discriminant analysis strictly requires the continuous independent variables (though dummy variables can be used as in multiple regressions).

There are two main uses of logistic regression: The first is the prediction of group membership. Since logistic regression calculates the probability of success over the probability of failure, the results of the analysis are in the form of an odds ratio. Logistic regression also provides knowledge of the relationships and strengths among the variables (Burns, R.P. and Burns, R., 2008). While the logistic regression model is nonlinear for probabilities (due to properties of its variation between 0 and 1), it is linear with respect to the logit coefficients. When we want to look at a dependence structure, with a dependent variable and a set of explanatory variables (one or more), we can use the logistic regression framework and when we have a proportion as a response, we use a logistic or logit transformation to link the dependent variable to the set of explanatory variables (Tranmer, & Elliot, 2008).

3.3.2 Logistic Regression Model

According to (Carter, & Signorino, 2010), suppose that we have k independent observations$y_1, \ldots, y_k$ and that the i^{th} observation can be treated as a realization of a random variable Y_i. We assume that Y_i has a binomial distribution:

$$Y_i \sim B(ni, pi) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.1).$$

with binomial denominator n_i and probability π_i. With individual data $n_i = 1$ for all i. This defines the stochastic structure of the model. Suppose further that the *logit* of the underlying probability p_i is a linear function of the predictor's logit $(p) = X_i{}'$ B, Odds of an event is the ratio of the probability that an event will occur to the probability that it will not occur. If the probability of an event occurring is p, the probability of the event not occurring is $(1\text{-}p)$. Then the corresponding odds is a value given by: odds of {Event} = $\frac{p}{1-p}$

Odds ratio: Another way out to interpret the coefficients is by considering only the ratio of the probability that the event occurs and the probability that the event does not occur (Gerolimetto, n.a). This ratio is:

$$\frac{\mathrm{p(Y=1)}}{\mathrm{p(Y=0)}} = \frac{\frac{e^{XB}}{1+e^{XB}}}{\frac{1}{1+e^{XB}}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.2)$$

In this case the exponentiated coefficients reflect **changes in the odds ratio**, consequently to a unit variation in the explicative variable. Coefficients (β) are in particular useful to determine the sign of the relationship: a positive coefficient indicates that a unit increase in the X is connected with increases the predicted probability and vice versa.

The logistic regression solution to this problem is to transform the odds using the natural logarithm (Peng, Lee & Ingersoll, 2002). With logistic regression we model the natural log odds as a linear function of the explanatory variables:

$$\text{logit (y)= ln (odds)= ln } \left(\frac{p}{1-p}\right) = a + B_1X_1 + B_2X_2 + \ldots + B_kX_k \text{..(3.3)}$$

Where p is the probability of interested outcome and x_k is the explanatory variables. The parameters of the logistic regression are α and B_k.

$$\mathrm{P} = \frac{e^{a+B_1X_1+B_2X_2+\ldots+B_kX_k}}{1-e^{a+B_1X_1+B_2X_2+\ldots+B_kX_k}} = \frac{1}{1-e^{-(a+B_1X_1+B_2X_2+\ldots+B_kX_k)}} \ldots \quad (3.4)$$

Where p probability that a case is in a particular category, e base of natural logarithms (approx 2.72), α constant of the equation and, B coefficient of the predictor variables.

This ought to look somewhat similar to the log odds equation. The odds ratio for a particular predictor variable is defined as e^{β}, where β is the logit coefficient estimate for the predictor and *e* is the natural log. If β is zero, the odds ratio will equal 1 (i.e., since any number to the 0 power is 1), which leaves the odds unchanged. If β is positive, the odds ratio will be greater than 1, which means the odds are increased. If β is negative, the coefficient will be less than 1, which means the odds are decreased (Ron Heck, 2012).

3.3.3 Multinomial Logistic Regressions

When the dependent variable has more than two values, there will be more than one regression equation. In fact, the number of regression equations is equal to one less than the number of outcomes. This makes interpretation more difficult because there is several regression coefficients associated with each independent variable.

In this case, care must be taken to understand what each regression equation is predicting. Although the type of data used for the dependent variable is different from that of multiple regressions, the practical use of the procedure is similar (Myers, Montgomery, Vining, and Robinson, 2012).

Multiple regressions can also be incorporated into the logistic regression model as well. Suppose we have *p* regress or variables $x1$, $x2, \ldots, xp$. Then we can generalize and define a multiple logistic regression function:

$$p(x1, \ldots, xp) = \frac{\exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p)} \quad \ldots\ldots\ldots\ldots\ldots (3.5)$$

And the logit of $p(x)$ is logit $(p(x)) = \beta_0 + \beta_1 x_1 + \cdot\cdot\cdot + \beta_p x_p$. Which shows that logistic regression is really just a standard linear regression model, once we transform the dichotomous outcome by the logit transform. Maximum likelihood is generally used to estimate the β_j's and their standard errors for the multiple logistic regression models as was done for the simple logistic regression.

3.3.4 Logit Model

According to (Torres-Reyna, 2012) use logit models whenever your dependent variable is binary (also called dummy) which takes values 0 or 1.

Logit regression is a nonlinear regression model that forces the output (predicted values) to be either 0 or 1. Logit models estimate the probability of your dependent variable to be 1 (Y=1). This is the probability that some event happens. The logit model is:

$$p_r\ (Y = 1) = F(B_0 + B_1X_1 + B_2X_2 \dots + \ B_kX_k) \dots \dots \dots \dots \dots \dots (3.6)$$

$$\text{Pr (Y=1)} = \frac{1}{1-e^{-((B_0 + B_1X_1 + B_2X_2\ + \dots + B_kX_k)}} \ \dots\dots\dots\dots\dots \quad (3.7)$$

$$\text{Pr (Y=1)} \quad = \frac{1}{1+\left[\frac{1}{e^{-((B_0 + B_1X_1 + B_2X_2\ + \dots + B_kX_k)}}\right]}$$

Where I = $B_1 + B_2X_2 \dots + \ B_kX_k$ is an index function, with the restriction that ∂P/∂I > 0. However, the logit model assumes that F is a logistic cumulative distribution function. Thus, the conditional mean function for the logit model is given by:

$$P = \frac{1}{1 + \exp(-I)} \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots (3.8)$$

Logit and probit models are basically the same; the difference is in the distribution:

- Logit – Cumulative standard logistic distribution (F)
- Probit – Cumulative standard normal distribution (Φ)

Both models provide similar results.

Like the probit model, the logit model assumes that the conditional mean function is given by:

P = F (I) = F($\beta_1 + \beta_2X_2 + \dots + \beta_kX_k$) …………………..(3.9)

3.4 Binary Logistic Regression Model

According to (Bolin, 2014).Binary logistic regression is a form of regression which is used when the dependent variable is a true or false dichotomy and the independent variables are of any type.

Logistic regression applies maximum likelihood estimation after transforming the dependent into a logit variable. A logit is the natural log of the odds of the dependent equaling a certain value or not (usually 1 in binary logistic models, or the highest value in multinomial models). Logistic regression estimates the odds of a certain event (value) occurring. This means that logistic regression calculates changes in the log odds of the dependent, not changes in the dependent itself as does OLS regression.

Logistic regression has many analogies to OLS regression: logit coefficients correspond to b coefficients in the logistic regression equation; the standardized logit coefficients correspond to beta weights, and a pseudo R^2 statistic is available to summarize the overall strength of the model. Unlike OLS regression, however, logistic regression does not assume linearity of the relationship between the raw values of the independent variables and raw values of the dependent; does not require normally distributed variables; does not assume homoscedasticity; and in general has less stringent requirements.

Logistic regression does, however, require that observations be independent and that the independent variables be linearly related to the logit of the dependent. The predictive success of logistic regression can be assessed by looking at the classification table, showing correct and incorrect classifications of the dichotomous, ordinal, or polytomous dependent variable. Goodness-of-fit tests such as the likelihood ratio test are available as indicators of model appropriateness, as is the Wald statistic to test the significance of individual independent variables.

Logistic regression is generally thought of as a method for modeling in situations for which there is a binary response variable. The predictor variables can be numerical or categorical (including binary), (Bukhari, n.a). But here the focus will be on the typical binary response version.

The logistic curve is better for modeling binary dependent variables coded 0 or 1 because it comes closer to hugging the y=0 and y=1pointson the y axis. Even more, the logistic function is bounded by 0 and 1, whereas the OLS regression functionmaypredictvaluesabove1andbelow 0.

$$P = \frac{e^{x'B}}{1+e^{x'B}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(3.10)}$$

3.4.1 Model Specification

We use a binary logistic regression model given that the dependent variable is dichotomous: 0 when a household is above and 1 when below the poverty line. Predictor variables are a set of socioeconomic and demographic status indicators and human capital and dwelling endowment of the household. They contain both dichotomous and continuous variables.

Logistic regression does not assume a linear relationship between the dependent and independent variables, the dependent variables do not need to be normally distributed, there is no homogeneity of variance assumption, in other words, the variances do not have to be the same within categories, normally distributed error terms are not assumed and the independent variables do not have to be interval or unbounded (Wright, 1995). Linear regression might be used as a preliminary step with a binary dependent variable to identify explanatory variables that are good predictors of the dependent variable, particularly if the software packages available to the analyst have variable selection procedures built into the linear regression software but not into the logistic regression software.

Let P_j denote the probability that the j-th household is below the poverty line. We assume that P_j is a Bernoulli variable and its distribution depends on the vector of predictors X, so that:

$$P_j(x) = \frac{e^{\alpha+Bx}}{1+e^{\alpha+Bx}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \text{(3.11)}$$

Where, β is a row vector and α a scalar. The logit function to be estimated is then written as:

$$\ln\frac{p_j}{1-p_j} = \alpha + \sum B_i x_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad \text{(3.12)}$$

The logit variable $\ln\{P_j/(1-P_j)\}$is the natural log of the odds in favor of the household falling below the poverty line. Equation 2 is estimated by maximum likelihood method and the procedure does not require assumptions of normality or homoscedasticity of errors in predictor variables.

Therefore, ln (P/1-P) = 1, if the household is poor while ln (P/1-P) = 0, if otherwise i.e non-poor.

Implicitly, the model is empirically estimated as $Y_i = B_0 + B_1X_{1i} + ... + B_3X_{3i} +......B_kX_{ki} +u_i$ --- (3.13)
Where i = 1, 2, 3, …, n.

In this multiple regression equation model, Yi is dependent variable (Poverty status; 1 = poor and 0 = non- poor) and X_1, X_2, … ,X_k are independent explanatory variables. $B_1$is the intercept, shows the average value of Y, when X_1, X_2, … , X_k are set equal to zero; B_0, B_1, …, B_k are partial regression/slope coefficients; u_i is the stochastic disturbance term; i is the i^{th} observation.

Testing the Joint Significance of All Predictors:

Test $H_0 : \beta_1 = \beta_2 = ... = 0$ versus the alternative that at least one of the coefficients $\beta_1, . . . , \beta_k$ is not zero.

This is like the overall F−test in linear regression. In other words, this is testing the null hypothesis that an intercept-only model is correct

Following is the multiple regression model specification:

$\text{Log}(p_i) = \ln\left[\frac{p}{1-p}\right] = B_0 + B_1\text{PR} + B_2\text{HS} + B_3\text{SHH} ...+ B_{12}\text{Msl} + \varepsilon_{13}$

The main statistical analysis applied in this study is logistic regression analyses. First, households were grouped into poor and non-poor households using the poverty line was calculated to be 2,966 SDG per.

3.4.2 The Logit and Logistic Transformations

According to (Hosmer, et al, 2013), in multiple regressions, a mathematical model of a set of explanatory variables is used to predict the mean of a continuous dependent variable. In logistic regression, a mathematical model of a set of explanatory variables is used to predict a logit transformation of the dependent variable.

Suppose the numerical values of 0 and 1 are assigned to the two outcomes of a binary variable. Often, the 0 represents a negative response and the 1 represents a positive response. The mean of this variable will be the proportion of positive responses. If p is the proportion of observations with an outcome of 1, then 1-p is the probability of an outcome of 0.

The ratio p/ (1-p) are called the odds and the logit is the logarithm of the odds, or just log odds. Mathematically, the logit transformation is written:

$$l = logit(p) = ln\left[\frac{p}{1-p}\right] \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \; (3.14)$$

The logistic equation logistic regression centers on the following terms:

Odds: An odd is a ratio formed by the probability that an event occurs divided by the probability that the event does not occur.

In binary logistic regression, the odd is usually the probability of getting a "1" divided by the probability of getting a "0". That is, in binary logistic regression,"1"is predicted and "0" is usually the reference category.

Odds ratio: Is a measure of effect size, describing the strength of association or non independence between two binary data values. It treats the two variables being compared symmetrically and can be estimated using some type of non-random samples. It is used as a descriptive statistic and plays an important role in logistic regression. An odds ratio is the ratio of two odds, such as the ratio of the odds for men and the odds for women. Here $p/ (1 - p)$ measures the probability that $y = 1$ relative to the probability that $y = 0$ and is called the **odds ratio** or **relative risk**. Odds ratios are the main effect size measure for logistic regression, reflecting in this case what difference gender makes as a predictor of some dependent variable. An odds ratio of 1.0 (which is 1:1 odds) indicates the variable has no effect, the furtherfrom1.0 in either direction, the greater the effect.

3.4.3 The Logistic Regression and Logit Models

In logistic regression, a categorical dependent variable Y having G (usually G =2) a unique value is regressed on a set of p independent variables X_1, X_2… X_p. For example, Y may be the presence or absence of a disease, condition after surgery, or marital status. Since the names of these partitions are arbitrary, we often refer to them by consecutive numbers. That is, in the discussion below, Y will take on the values 1, 2, … G.Let

$$X = (X_1, X_2, \cdots, X_p)$$

$$B_g = \begin{pmatrix} \beta_{g1} \\ \vdots \\ \beta_{gp} \end{pmatrix}$$

The logistic regression model is given by the G equations

$$\begin{aligned} \ln\left(\frac{p_g}{p_1}\right) &= \ln\left(\frac{P_g}{P_1}\right) + \beta_{g1}X_1 + \beta_{g2}X_2 + \cdots + \beta_{gp}X_p \\ &= \ln\left(\frac{P_g}{P_1}\right) + XB_g \end{aligned} \quad \text{………..… (3.15)}$$

Here, p_g is the probability that an individual with values X_1, X_2,...,X_p is in outcome g. That is, p_g = Pr (Y = g | X). Usually $X_1 \equiv 1$ (that is, an intercept is included), but this is not necessary. The quantities P_1, P_2, ..., P_G represent the prior probabilities of outcome membership. If these prior probabilities are assumed equal, then the term $\ln(P_g/ P_1)$ becomes zero and drops out. If the priors are not assumed equal, they change the values of the intercepts in the logistic regression equation. Outcome one is called the reference value. The regression coefficients β_{11}, β_{12},…, β_{1p} for the reference value are set to zero. The choice of the reference value is arbitrary. Usually, it is the most frequent value or a control outcome to which the other outcomes are to be compared. This leaves G-1 logistic regression equations in the logistic model. The β's are population regression coefficients that are to be estimated from the data. Their estimates are represented by b's. The β's represents unknown parameters to be estimated, while the b's are their estimates. These equations are linear in the logit of p. However, in terms of the probabilities, they are nonlinear. The corresponding nonlinear equations are

$$p_g = \text{Prob}(Y = g \mid X) = \frac{e^{XB_g}}{1 + e^{XB_2} + e^{XB_3} + \cdots + e^{XB_G}} \quad \text{………….. (3.16)}$$

Since $e^{xB} = 1$ because all of its regression coefficients are zero, a note on the names of the models. Often, all of these models are referred to as logistic regression models.

However, when the independent variables are coded as ANOVA type models, they are sometimes called logit models.

A note about the interpretation of e^{xB} may be useful. Using the fact that x^{a+b} = (e^a) (e^b), e^{xB} may be re-expressed as follows:

$$e^{XB} = e^{B_1X_1} + e^{B_2X_2} + \cdots + e^{B_pX_p}$$

$$= e^{B_1X_1} e^{B_2X_2} \ldots e^{B_pX_p} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.17)$$

This shows that the final value is the product of its individual terms (Hosmer, Lemeshow, and Sturdivant, 2013).

Logit: The "logit function" is the function used in logistic regression to transform the dependent variable prior to attempting to predict it. Specifically, the logit function in logistic regression is the log odds, explained above. The "logit" is the predicted value of the dependent variable. "Logit coefficients" are the b coefficients in the logistic equation. For the logit model $\ln\frac{p}{1-p} = x'B$

Parameter estimates: These are the logistic (logit or b) regression coefficients for the independent variables and the constant in a logistic regression equation, much like the b coefficients in OLS regression. Synonyms for parameter estimates are un-standardized logistic regression coefficients, logit coefficients, log odds-ratios, and effect coefficients. Parameter estimates are on the right-hand side of the logistic regression equation and logit is on the left-hand side, used to arrive at the predicted value.

Let P_i denotes the probability that the i^{th} household is below the poverty line. We assume that the P_i is a Bernoulli variable and its distribution depends on the vector of predictors X, so

$$P_i(X) = \frac{e^{\beta X}}{1+e^{\beta X}} \quad \ldots\ldots(3.18)$$

where β is a row vector

The logit function to be estimated is then written as

$$ln\frac{P_i}{1-P_i} = \sum_{j=1}^{k} \beta_j X_{ij} \quad .. \ . \ldots (3.19)$$

$ln\frac{P_j}{1-P_j}$ is the natural log of the odds in favor of the household falling below the poverty line whereas β_j is the measure of change in the logarithm of the odds ratio of the chance of the poor to non poor household and can also be written as

$$\frac{\partial \log(oddratio)}{\partial X_j} = -\beta_j$$

The marginal effects are also computed that show the change in the probability when there is a unit change in the independent variables. The marginal effects are computed as follows:

$$\frac{\partial P}{\partial X_j} = \frac{B_j e^{-z}}{[1+e^{-z}]^2} where z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \ldots + \beta_k X_k \quad (3.20)$$

Finally we summarize for multiple regression a model of the following form can be used to predict the value of a response variable y using the values of a number of explanatory variables:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \ldots + \beta_k X_k \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \; (3.21)$$

B_0 = constant / intercept, $B_1 \rightarrow B_k$ coefficient for k explanatory variables $x_1 \rightarrow x_k$

The regression process finds the co-efficient which minimize the squared differences between the observed and expected values of y (the residuals). As the outcome of logistic regression is binary, y needs to be transformed so that the regression process can be used. The logit transformation gives the following:

$$\ln\left[\frac{p}{1-p}\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \ldots + \beta_k X_k \ldots\ldots\ldots\ldots\ldots \; (3.22)$$

p= probability of uneven occurring, $\frac{p}{1-p}$ adds ratio

If the probabilities of the event of interest happening for individuals are needed, the logistic regression equation can be written as:

$$P = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \ldots + \beta_k X_k)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \ldots + \beta_k X_k)}, \; 0 < p < 1 \; \ldots\ldots\ldots\ldots \; (3.23)$$

3.4.4 Assumptions of the Binary Logistic Regression Model

Logistic regression does not make many of the key assumptions of linear regression and general linear models that are based on ordinary least squares algorithms – particularly regarding linearity, normality, homoscedasticity, and measurement level.

Firstly, it does not need a linear relationship between the dependent and independent variables. Logistic regression can handle all sorts of relationships, because it applies a non-linear log transformation to the predicted odds ratio.

Secondly, the independent variables do not need to be multivariate normal – although multivariate normality yields a more stable solution. Also the error terms (the residuals) do not need to be multivariate normally distributed. Thirdly, homoscedasticity is not needed.

Logistic regression does not need variances to be heteroscedastic for each level of the independent variables. Lastly, it can handle ordinal and nominal data as independent variables. The independent variables do not need to be metric (interval or ratio scaled).

3.5 Model Selection Methods

Method selection allows you to specify how independent variables are entered into the analysis. Using different methods, you can construct a variety of regression models from the same set of variables.

According to (HosmerJr, et al, 2013), in addition to All Effects, different techniques for the automatic model building are available for logistic regression. Specifically, forward stepwise, backward stepwise, forward entry, backward removal, and best-subset search procedures are available in STATISTICA and are described below. Stepwise Logistic Regression Stepwise Logistic Regression methods, specifically the Forward Stepwise and Backward Stepwise methods, are used to perform a stepwise selection of predictor variables.

During the forward step of stepwise model building, if two or more effects have p-values that are so small as to be virtually indistinguishable from 0, STATISTICA will select the effect with the largest score statistic if the degrees of freedom for all effects in question are equal. If the effects differ with respect to the degrees of freedom, the Score statistics are normalized using the Wilson-Hilferty transformation, and the effect with the largest transformed value is entered into the model. For the backward step, if the *p*-values for two or more effects are virtually indistinguishable from 1, STATISTICA will remove the effect with the smallest Wald statistic in the case of equal degrees of freedom and the smallest normalized value in the case of unequal degrees of freedom.

• **Enter** a procedure for variable selection in which all variables in a block are entered in a single step.

• **Forward Selection (Conditional)**. Stepwise selection method with entry testing based on the significance of the score statistic, and removal testing based on the probability of a likelihood-ratio statistic based on conditional parameter estimates.

• **Forward Selection (Likelihood Ratio)**. Stepwise selection method with entry testing based on the significance of the score statistic, and removal testing based on the probability of a likelihood-ratio statistic based on the maximum partial likelihood estimates.

• **Forward Selection (Wald)**. Stepwise selection method with entry testing based on the significance of the score statistic, and removal testing based on the probability of the Wald statistic.

3.6 Models Building Strategy

Automatic stepwise selection procedure:

• Start with a list of important covariates obtained as before using the univariate analysis.

• Forward selection: Start with a simple model and add terms sequentially until further additions do not significantly improve the fit.

• Backward elimination: Start with a complex model and remove terms sequentially until a further deletion leads to a significantly poorer fit (Generally preferred over forward selection).

•Other variants.

•Cannot trust the results.

• Can also use a penalized measure of model fit such as Akaike Information Criterion (AIC), Adjusted R^2 and Bayesian Information Criterion (BIC) instead of p-values.

AIC = – 2(maximized log likelihood – # parameters in the model). Lower is better.

3.6.1 Forward Selection

Forward stepwise procedures: start from a simple ‘null’ model, and incrementally update fit to allow slightly more complexity.

It is better than backwards methods: the ‘full’ model can be expensive or tough to fit, while the null model is usually available in closed form. Jitter the data and the full model can change dramatically (because it is over fit). The null model is always the same. Stepwise approaches are ‘greedy’: they find the best solution at each step without thought to global path properties.

The method of forward selection proceeds as follows.

- Begin with no terms in the model.
- Find the term that, when added to the model, achieves the largest value of the log likelihood. Enter this term into the model.
- Continue adding terms until a target value for the log-likelihood is achieved or until a preset limit on the maximum number of terms in the model is reached. Note that these terms can be limited to those keeping the model hierarchical. And also you stop when some model selection rule (AIC) is lower for the current model than for any of the models that add one variable.

This method is comparatively fast, but it does not guarantee that the best model is found except for the first step when it finds the best single term. You might use it when you have a large number of observations and terms so that other, more time consuming, methods are not feasible.

3.6.2 Wilson-Hilferty Transformation

The Wilson-Hilferty transformation method transforms a χ^2 variable to the *Z*-scale so that their *p*-values are closely approximated. This transformation, therefore, enables the comparison of the statistical significance of the χ^2 values with different degrees of freedom.

The transformation is given by

$$W(y) = \frac{\left(\frac{Y}{n}\right)^{1/3} - \left(1-\left(\frac{1}{9}\right)\right)\left(\frac{2}{n}\right)}{\sqrt{\left(\frac{1}{9}\right)\left(\frac{2}{n}\right)}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (3.24)$$

Where,

Y = x^2 statistic

n = Degrees of freedom

3.7 Estimation Binary Response Probabilities

According to (Stephenson, 2008) we examine the analysis of binary response data. Binary response data abounds in many application areas and presents a unique problem because ordinary least squares simple linear regression is an inappropriate means of analysis. Performing simple linear regression on the logit-transformed data corrects for the non-linear nature of the binary response but does not address the violation of equal variance and normality assumptions. The use of maximum likelihood estimation provides a means of working with binary response data.

Many dependent variables of interest in economics and other social sciences can only take two values. The two possible outcomes are usually denoted by 0 and 1. Such variables are called dummy variables or dichotomous variables (Blundell, & Powell, 2004).

The expected value of a dichotomous variable $y_i \in \{0, 1\}$ is the probability that it takes the value 1:

$E(y_i) = 0 \cdot P(y_i = 0) + 1 \cdot P(y_i = 1) = P(y_i = 1)$ ………………..... *(3.25)*

The linear regression model,

$y_i = x_0\ i_B + v_i,\ E(v_i|x_i) = 0$ ………………….(3.26)

Is called linear probability model in this context, this linear model is not an adequate statistical model as the expected value $E\ (y_i|x_i) = x_0\ i_B$ can lie outside [0, 1] and does not represent a probability. In addition, the error term is heteroskedasticity as $V\ (v_i|x_i) = x_0\ i_B(1 - x_0\ i_B)$ depends on x_i.

The Bernoulli distribution: We have already encountered a distribution for outcomes which take on only two values – the Bernoulli distribution

$f(1) = \pi$

$f(0) = 1 - \pi$

Where the event occurs with probability π and fails to occur with probability $1 - \pi$. Recall that the likelihood of the Bernoulli distribution is

$$l = \prod_{i=1}^{N} \pi^{yi} (1-\pi)^{1-yi} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.27)$$

And that the log-likelihood is

$$\ln\mathcal{L} = \sum_{i=1}^{N} [y_i \ln(\pi) + (1-y_i)\ln(1-\pi)]$$

The Bernoulli distribution would be an appropriate model of dichotomous choice if each event had the same chance of occurring. However, it is not a good model for widely variable outcomes. Thus, the Bernoulli distribution is too restrictive as it stands. Instead we would like to let π vary across cases i.e. π_i. This keeps the Bernoulli form but allows us to capture variation across cases in the probability. However, we can't just let each observation have its own π_i since the model would not be identified.

This is why we write $\pi_i = g(X, \beta)$ in order to both reduce the number of parameters and to add substantive explanatory variables. In effect, we have Y_i *bern*$\sim$ (πi) (8) where $\pi_i = g\ (x_i, \beta)$. All we need now is to find a function $g\ (\cdot)$ of the x_s and the β_s to substitute into the Bernoulli likelihood function i.e.

$$\mathcal{L} = \prod_{i=1}^{N} g(x_i, \beta)^{y_i} (1 - g(x_i, \beta))^{1-y_i} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.28)$$

Or, alternatively, the log-likelihood function i.e. the log-likelihood for binary data with a logistic model is given by

$$\ln\mathcal{L} = \sum_{i=1}^{N} \{y_i \ln[g(x_i, \beta)] + (1-y_i)\ln[(1 - g(x_i, \beta)]\} \quad ..(3.29)$$

3.8 Estimating of the Model Parameters

In this section we shall discuss how model parameters are estimated using the method of maximum likelihood and Pseudo R^2 Measures and, assessment of the fitted model using a Wald $\chi 2$ statistic, the likelihood ratio test, Deviance Test, Hosmer - Lemeshow Test and Sc.

3.8.1 Maximum Likelihood Estimation

The maximum likelihood estimate is that the value of the parameter that makes the observed data most likely. The logistic regression model just developed is a generalized linear model with binomial errors and link logit. We can, therefore, rely on the general theory developed in logistic regression to obtain estimates of the parameters and to test hypotheses.

Small (Finite) sample tests, e.g., t-test and F-test, cannot be used to test hypotheses in the linear probability model. This is because the error term has a binomial distribution, not a normal distribution. To test hypotheses, you must use large sample (asymptotic) tests. These include the t-test, approximate F-test, Likelihood ratio test, Wald test, and Hosmer - Lemeshow Test. MLE allows more flexibility in the data and analysis because it has fewer restrictions (HosmerJr, et al, 2013).

We want to choose β's that maximizes the probability of observing the data we have:

$$L = \Pr(y_1, y_2, \ldots, y_N) = \Pr(y_1)\Pr(y_2)\ldots\Pr(y_N) = \prod_{i=1}^{N} \Pr(y_i) \quad \ldots..(3.30)$$

Substituting in using logistic regression model:

$$\ln L = \sum_i y_i \beta x_i - \sum_i \ln\left(1 + \exp\left(\beta x_i\right)\right)$$

If the value is less than (0.05), which confirms the significance of the model The maximization of the likelihood is achieved by an iterative method called Fisher scoring. Fisher scoring is similar to the Newton-Raphson procedure except that the hessian matrix (matrix of second order partial derivatives) is replaced with its expected value (Longford 1994). The Fisher scoring update formula for the regression coefficients is given by

$$\hat{B}_{k+1} = \hat{B}_k + [1(\hat{B}_k)]^{-1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.31)$$

Where, $\hat{B}_k$ = estimate of B based on k^{th} iteration

The algorithm completes when the convergence criterion is satisfied or when the maximum number of iterations has been reached. Convergence is obtained when the difference between the log-likelihood function from one iteration to the next is small. By default, the convergence criterion is 1e-7, and thus convergence is obtained when $|\ln L_{k+1} - \ln L_k| \leq le - 7$ ……………………………. (3.32)

Where, L_k = likelihood evaluated at $\hat{B}_k$

In the equations below:

$\hat{\pi}_i =$ Estimated probability of i^{th} case

The score vector is given by:

$$s(\hat{B}) = \frac{\partial 1}{\partial \hat{B}} \begin{bmatrix} \sum_{i=1}^{N} \omega_i (y_i - \hat{\pi}) \\ \sum_{i=1}^{N} \omega_i (y_i - \hat{\pi}) X_{i1} \\ . \\ . \\ \sum_{i=1}^{N} \omega_i (y_i - \hat{\pi}) X_{ip} \end{bmatrix} \quad \text{…………………. (3.33)}$$

The information matrix is given by:

$$1(\hat{B}) = -E\left[\frac{\partial^2 1}{\partial \hat{B}_i \partial \hat{B}_j}\right] \begin{bmatrix} \sum_{i=1}^{N} \omega_i (\hat{\pi}_i)(1-\hat{\pi}_i) \ldots \sum_{i=1}^{N} \omega_i (\hat{\pi}_i)(1-\hat{\pi}_i) X_{ip} \\ \sum_{i=1}^{N} \omega_i (\hat{\pi}_i)(1-\hat{\pi}_i) X_{i1} \ldots \sum_{i=1}^{N} \omega_i (\hat{\pi}_i)(1-\hat{\pi}_i) X_{i1} X_{ip} \\ \vdots \\ . \\ \sum_{i=1}^{N} \omega_i (\hat{\pi}_i)(1-\hat{\pi}_i) X_{ip} \ldots \sum_{i=1}^{N} \omega_i (\hat{\pi}_i)(1-\hat{\pi}_i) X_{ip} X_{ip} \end{bmatrix}$$

The asymptotic estimated covariance matrix of the estimated coefficients is given by

$$\hat{\Sigma}_{\hat{\beta}} = I(\hat{\boldsymbol{\beta}})^{-1}$$

Testing the overall model:

$$H_0 : \beta_1 = \cdots = \beta_k = 0$$
$$H_A : \text{Not all } \beta_i = 0$$
$$T.S.\ X^2_{obs} = (-2\log(L_0)) - (-2\log(L_1))$$
$$R.R.\ X^2_{obs} \geq \chi^2_{\alpha,k}$$
$$P = P(\chi^2 \geq X^2_{obs})$$

L_0, L_1 are values of the maximized likelihood function, computed by statistical software packages. This logic can also be used to compare full and reduced models based on subsets of predictors. Testing for individual terms is done as in model with a single predictor.

3.8.2 Pseudo R^2 Measures

There is not an easily defined R^2 with the logistic regression that can be used to quantify the variance accounted for in the response variable. There are, however, a number of pseudo-R^2 values that have been proposed based on improvement in fit (reduction in deviance) when one or more variables are added to the model. The Pseudo-R^2 in logistic regression is best used to compare different specifications of the same model. Don't try to compare models with different data sets with the Pseudo-R^2. The most common are the Cox and Snell (Cox & Snell, 1989; Cragg & Uhler, 1970; Maddala, 1983) and Nagelkerke (1991) pseudo R2 values. Each has values that theoretically range between 0 and 1.

3.8.2.1 Cox-Snell R^2

In linear regression using ordinary least squares, a measure of goodness of fit is R^2, which represents the proportion of variance explained by the model. Using logistic regression, an equivalent statistic does not exist, and therefore several pseudo-R^2statistics have been developed. The *Cox-Snell* R^2 is a pseudo - R^2statistic, and the ratio of the likelihoods reflects the improvement of the full model over the intercept-only model with a smaller ratio reflecting greater improvement (HosmerJr, D.W., Lemeshow, S. and Sturdivant, R.X., 2013). It is given by:

$$\text{Cox-snell}R^2 = 1-\left[\frac{L(R)}{L(F)}\right]^{2/N} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.34)$$

Where, L(R) = likelihood of intercept-only model, L(F) = likelihood of the specified model, N = Number of observations.

3.8.2.2 NagelkerkeR^2

The Nagelkerke R^2 adjusts the Cox-Snell R^2so the range of possible values extends to one (HosmerJr, Lemeshow, et al, 2013).

Nagelkerke $R^2 = \frac{1-\left[\frac{L(R)}{L(F)}\right]^{2/N}}{1-L(R)^{2/N}}$ …………………………. (3.35)

Where, L(R) = likelihood of intercept-only model, L(F) = likelihood of specified model, N = Number of observations.

3.8.3 Likelihood Ratio Test

The likelihood ratio test is used to compare the fit of two models, one of which is nested within the other. This is typically performed to determine if a simpler model can be used to adequately model the data. The test is based on a comparison of full and reduced models where both models are fitted to the data and their log-likelihoods are calculated. Let the full model (F) have *p* parameters and the reduced model (R) have *q* parameters such that $q < p$ (HosmerJr, Lemeshow et al, 2013).

Full Model

Logit $[\pi(x)]= \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = B_0+B_1X_1+\ldots+B_{q-2}X_{q-2}+B_{q-1}X_{q-1}+B_qX_q+B_{q+1}+\ldots+B_{p-1}$

.... …………………………………….. (3.36)

Reduced Model

Logit $[\pi(x)] = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = B_0+B_1X_1+\ldots+B_{q-2}X_{q-2}+B_{q-1}X_{q-1}$ ..(3.37)

Let L (F) denote the maximized log-likelihood of the full model and L(R) represent the maximized log-likelihood of the reduced model. The null and alternative hypotheses with respect to this test are shown below.

H_0: $B_q= B_{q+1} = \ldots= B_{p-1}=0$

H_a: not H_0

The test statistic is given by: LR = -2[L(F)-L(R)]……….(3.38)

This LR statistic is asymptotically distributed as χ^2 with *p-q* degrees of freedom. Or the Likelihood Ratio Statistic:

$$G^2 = -2\log_e\left[\frac{L(R)}{L(F)}\right] = -2[\log_e L(R) - \log_e L(F)]$$

……………………..(3.39)

The Decision rule: If $G^2 \leq \chi^2(1-\alpha;p-q)$, conclude H_0, If $G^2 > \chi^2(1-\alpha;p-q)$, conclude H_a.

The LR is used to Comparison of null or constant only model to the full model which includes the predictors. Can be used to compare any two "nested" models (Homser, & Lemeshow, 1989), recommend against the use of RL2 as the goodness of fit measure. However, we have included it in our output because it does provide a comparative measure of the proportion of the log-likelihood that is accounted for by the model. Just remember that an RL 2 value of 1.0 indicates that the logistic regression model achieves the same log likelihood as the saturated model. However, this does not mean that it fits the data perfectly. Instead, it means that it fits the data as well as could be hoped for.

3.9 Goodness of Fit Statistics

After estimating the regression coefficients, it is necessary to assess the appropriateness, adequacy, and usefulness of the model. First the importance of each of the explanatory variables is assessed by carrying out Wald χ^2statistic or a likelihood ratio test. The overall goodness of fit of the model is then tested.

H_0 : the model fits well

H_A : the model does not fit well

According to (Read, & Cressie, 2012) Instead of using maximum likelihood estimation we could estimate the parameters by minimizing the weighted sum of squares:

$$S_w = \sum_{i=1}^{N} \frac{(y_i - n_i\pi_i)^2}{n_i\pi_i(1-\pi_i)} \qquad \text{…………….. …… (3.40)}$$

since $\mathrm{E}(Y_i) = n_i\pi_i$ and $\mathrm{var}(Y_i) = n_i\pi_i(1-\pi_i)$.

This is equivalent to minimizing the **Pearson chi-squared statistic**

$$\chi^2 = \sum \frac{(0-e)^2}{e} \text{...} (3.41)$$

Where o represents the observed frequencies, e represents the expected frequencies. The reason is that:

$$\begin{aligned} X^2 &= \sum_{i=1}^{N} \frac{(y_i - n_i\pi_i)^2}{n_i\pi_i} + \sum_{i=1}^{N} \frac{[(n_i - y_i) - n_i(1-\pi_i)]^2}{n_i(1-\pi_i)} \\ &= \sum_{i=1}^{N} \frac{(y_i - n_i\pi_i)^2}{n_i\pi_i(1-\pi_i)}(1 - \pi_i + \pi_i) = S_w. \end{aligned}$$

.

When $\chi^2$is evaluated at the estimated expected frequencies, the statistic

$$X^2 = \sum_{i=1}^{N} \frac{(y_i - n_i\widehat{\pi}_i)^2}{n_i\widehat{\pi}_i(1-\widehat{\pi}_i)}$$

Which is asymptotically equivalent to the deviances?

$$D = 2\sum_{i=1}^{N} \left[y_i \log\left(\frac{y_i}{n_i\widehat{\pi}_i}\right) + (n_i - y_i)\log\left(\frac{n_i - y_i}{n_i - n_i\widehat{\pi}_i}\right)\right]. \text{...........} (3.42)$$

The proof of the relationship between X^2 and D uses the Taylor series expansion of $s\log(s/t)$ about $s = t$, namely,

$$s\log\frac{s}{t} = (s-t) + \frac{1}{2}\frac{(s-t)^2}{t} + \ldots \quad .$$

Thus

$$\begin{aligned} D &= 2\sum_{i=1}^{N}\{(y_i - n_i\widehat{\pi}_i) + \frac{1}{2}\frac{(y_i - n_i\widehat{\pi}_i)^2}{n_i\widehat{\pi}_i} + [(n_i - y_i) - (n_i - n_i\widehat{\pi}_i)] \\ &\quad + \frac{1}{2}\frac{[(n_i - y_i) - (n_i - n_i\widehat{\pi}_i)]^2}{n_i - n_i\widehat{\pi}_i} + \ldots\} \\ &\cong \sum_{i=1}^{N} \frac{(y_i - n_i\widehat{\pi}_i)^2}{n_i\widehat{\pi}_i(1-\widehat{\pi}_i)} = X^2. \end{aligned}$$

The asymptotic distribution of D, under the hypothesis, that the model is correct, is $D \sim x^2(N - p)$, therefore approximately $x^2 \sim x^2(N - p)$. The choice between D and $X2$ depends on the adequacy of the approximation to the $x^2(N - p)$ distribution. There is some evidence to suggest that $x^2$is often better than D because D is unduly influenced by very small frequencies (Cressie and Read, 1989).

Both the approximations are likely to be poor, however, if the expected frequencies are too small (e.g., less than 1).

3.9.1Chi-Square Tests / Omnibus Tests

With logistic regression, instead of R^2 as the statistics for the overall fit of the linear regression model, deviance between observed values from the expected values is used. In addition, Omnibus test as a general name refers to an overall or a global test; other names include F-test or Chi-squared test. In linear regression, residuals can be defined as $y_i - \hat{y}_i$, where y_i is the observed dependent variable for the i_{th} subject, and $\hat{y}_i$ the corresponding prediction from the model. The same concept applies to logistic regression, where y_i is equal to either 1 or 0, and the corresponding prediction from the model is as:

$$\hat{y}_i = \frac{\exp(\alpha+\beta_1 X_{i1}+\cdots\ldots+\beta_k X_{ik})}{1+\exp(\alpha+\beta_1 X_{i1}+\cdots\ldots+\beta_k X_{ik})} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.43)$$

Chi-square test can be based on the residuals, $y_i - \hat{y}_i$ (Peng & So, 2002).A standardized residual can be defined as

$$r_i = \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i(1-\hat{y}_i)}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.44)$$

Where the standard deviation of the residuals is $\hat{y}_i$ $(1\text{-}\hat{y}i)$, one can then form a x^2 statistic as

$$x^2 = \sum_{i=1}^{n} r_i^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3.45)$$

H_0 : the model not significant

H_A : the model is significant

This statistic follows a x^2 distribution with $n-(k+1)$ degrees of freedom, so that *p*-values can be calculated.

We reject the null hypothesis if the p-value $<$ 0.05 that means the model is significant and represents the data well.

3.9.2 Wald Test

The Wald test will be familiar to those who use multiple regressions.

In multiple regressions, the common t-test for testing the significance of a particular regression coefficient is a Wald test. In logistic regression, the Wald test is calculated in the same manner (HosmerJr et al, 2013). The formula for the Wald statistic is:

$$\text{Wald} = \left(\frac{\hat{\beta}}{\hat{\sigma}_{\hat{\beta}_i}}\right)^2 \qquad \text{…………………………..(3.46)}$$

where $\hat{\sigma}_{\hat{B}_i}$ is an estimate of the standard error of b provided by the square root of the corresponding diagonal element of the covariance matrix, V(β˄). With large sample sizes, the distribution of z_j is closely approximated by the normal distribution. With small and moderate sample sizes, the normal approximation is described as 'adequate.' The Wald test is used in SPSS to test the statistical significance of individual regression coefficients.

The *Wald statistic* is asymptotically distributed as with x^2 degree of freedom. The estimated standard error of the i^{th} estimated coefficient, $\hat{\sigma}_{\hat{B}_i}$, is the square root of the i^{th} diagonal element of the estimated covariance matrix ,$\widehat{\sum}_{\hat{B}}$ that is, $\hat{\sigma}_{ii}$ .When β is *k*-dimensional and asymptotic is normal, the hypothesis test is given by the following quadratic form[4]:

Wald = $(\hat{B} - B_0)^{T} \widehat{\sum}_{\hat{B}}^{-1} (\hat{B} - B_0)$ …………………. (3,47)

Where:$\widehat{\sum}_{\hat{B}}$ estimated variance-covariance matrix of $\hat{B}$

This statistic is asymptotically distributed as x^2 with degrees of freedom equal to the number of parameters estimated for a given effect and can be used to test the hypothesis that all parameters estimated for a given effect are equal to 0.

In many statistical inference procedures, we have used Chi-square distribution based on the likelihood ratio, Score, or Wald test statistics.

The global Chi-square addresses the question "Is this model better than nothing?" The answer "yes" suggests the acceptance of the model (Wu, & Zhang, 2006).

Confidence intervals for the regression coefficients are based on the Wald statistics. The formula for the limits of a 100(1− α) % two-sided confidence interval is

$b_j \pm$ /z α / 2 / s_{bj}

Idea is to use large sample *Z* statistic from a single model to test:

$$H_0 : \beta_K = 0$$

$$\text{Here, } Z = \frac{\hat{\beta}_K}{SE_{\hat{\beta}_K}} \text{ where } Z \sim N(0,1)$$

Critical *Z* value for α=0.05 is 1.96 (two-sided), It is significant if the probability of p- value less thanα.

3.9.3 Hosmer-Lemeshow Test

To calculate the Hosmer-Lemeshow goodness of fit test, the data are sorted first in their increasing order of predicted probability. The observations are divided into groups using the setting for HL Groups, with a default of 10 groups. The groups are constructed based on the percentiles of the estimated probabilities. See Hosmer and Lemeshow (2000) for details. The Hosmer-Lemeshow statistic is distributed as a distribution with g-2 degrees of freedom given the null hypothesis (Hosmer et al, 2013).

This goodness-of-fit statistic is more robust than the traditional goodness-of-fit statistic used in logistic regression, particularly for models with continuous covariates and studies with small sample sizes. Can provide improved estimates of fit when the sample size is large, with small samples (with n < 400, according to Hosmer & Lemeshow, 2000), its use is not recommended. It is based on grouping cases into deciles of risk and comparing the observed probability with the expected probability within each decile. To calculate the Hosmer - Lemeshow goodness of fit test, the data are sorted first in their increasing order of predicted probability. The observations are divided into groups using the setting for HL Groups, with a default of 10 groups. The groups are constructed based on the percentiles of the estimated probabilities. Statistic is distributed asχ^2 a distribution with g-2 degrees of freedom given the null hypothesis. The Hosmer - Lemeshow goodness of fit test statistic is given by:

$$\chi^2_{HL} = \sum_{j=1}^{g} \frac{\left(o_j - N_j\bar{\pi}_j\right)^2}{N_j\bar{\pi}_j(1-N_j\bar{\pi}_j)} \text{ ....} \quad (3.48)$$

Where, g = Number of groups, N_j = Numbere of observations in the groups, o_j Number of responses in the jth group, $\bar{\pi}_j$ = Averages of the predicted probability for the j^{th} group.

$H_{0:}$ the model is represented the data well

H_A : the model does not represented the data well

If the (sig) is greater than (0.05) then this indicates the goodness of the whole conciliation model.
Where it compares the value of this test with the tabular value of the chi square, If the calculated value is less than or equal tabulated this means, that we accept the H_0 of that the model is well fit to the data, but if the calculated value is larger, it means rejecting the null hypothesis and accept the alternative hypothesis that non-conformity the model data.

3.9.4 Deviance Test

The deviance is basically a measure of how much-unexplained variation there is in our logistic regression model – the higher the value the less accurate the model. It compares the difference in probability between the predicted outcome and the actual outcome for each case and sums these differences together to provide a measure of the total error in the model.

When the full model in the likelihood ratio test statistic is the saturated model, LR is referred to as the deviance. A saturated model is one which includes all possible terms (including interactions) so that the predicted values from the model equal the original data. The formula for the deviance is

D = −2 [L Reduced – L Saturated]. The deviance may be calculated directly using the formula for the deviance residuals.

This formula is following, the deviance function, which is also used as a goodness-of-fit statistic for logistic models, is defined as:

$$D = 2\sum_{j=1}^{J}\sum_{g=1}^{G} w_{gj} \ln\left(\frac{w_{gj}}{n_j p_{gj}}\right) \text{ ....} (3.49)$$

This expression may be used to calculate the log likelihood of the saturated model without actually fitting a saturated model. The formula is

$$L_{\text{Saturated}} = L_{\text{Reduced}} + \frac{D}{2}$$

The deviance in logistic regression is analogous to the residual sum of squares in multiple regressions. In fact, when the deviance is calculated in multiple regressions, it is equal to the sum of the squared residuals. Deviance residuals, to be discussed later, may be squared and summed as an alternative way to calculate the deviance, D. The change in deviance, ΔD, due to excluding (or including) one or more variables is used in logistic regression just as the partial F test is used in multiple regressions. Many texts use the letter G to represent ΔD, but we have already used G to represent the number of groups in Y. Instead of using the F distribution, the distribution of the change in deviance is approximated by the chi-square distribution. Note that since the log likelihood for the saturated model is common to both deviance values, ΔD is calculated without actually estimating the saturated model. This fact becomes very important during subset selection. The formula for ΔD for testing the significance of the regression coefficient(s) associated with the independent variable$X1$is:

$$\begin{aligned}\Delta D_{X1} &= D_{\text{without}\,X1} - D_{\text{with}\,X1} \\ &= -2\left[L_{\text{without } X1} - L_{\text{Saturated}}\right] + 2\left[L_{\text{with } X1} - L_{\text{Saturated}}\right] \\ &= -2\left[L_{\text{without } X1} - L_{\text{with } X1}\right]\end{aligned}$$

Note that this formula looks identical to the likelihood ratio statistic. Because of the similarity between the change in deviance test and the likelihood ratio test, their names are often used interchangeably. Can be computed for any model, distributed as chi-square value.

3.10 Model Validation

Logistic regression models are frequently used to predict a dependent variable from a set of independent variables. An important question is whether the results of the logistic regression analysis on the sample can be extended to the population the sample has been chosen from. This question is referred to as model validation. In practice, a model cab is validated by deriving a model and estimating its coefficients in one data set, and then using this model to predict the outcome variable from the second data set, then checks the residuals, and so on. When a model is validated using the data on which the model was developed, it is likely to be over- estimated. Thus, the validity of model should be assessed by carrying out tests of goodness of fit and discrimination on a different data set (Giancristofaro & Salmaso, 2007). If the model is developed with a sub sample of observations and validated with the remaining sample, it is called internal validation. The most widely used methods for obtaining a good internal validation are data-splitting, repeated data-splitting, jackknife technique and bootstrapping (Harrell, Lee, & Mark, 1996).

If the validity is tested with a new independent data set from the same population or from a similar population, it is called external validation. Obtaining a new data set allows us to check the model in a different context. If the first model fits the second dataset, there is some assurance of generalization ability of the model. However, if the model does not fit the second data, the lack of fit can be either due to the different contexts of the two data sets, or true lack of fit of the first model (Park, 2013).

3.10.1 Classification Tables

The classification table is a method to evaluate the predictive accuracy of the logistic regression model (Peng & So, 2002). In this table the observed values for the dependent outcome and the predicted values (at a user-defined cut-off value) are cross-classified. For example, if a cutoff value is 0.5; all predicted values above 0.5 can be classified as predicting an event, and all below 0.5 as not predicting the event.

Then two-by-two tables of data can be constructed with dichotomous observed outcomes, and dichotomous predicted outcomes. The table has following form (Park, H., 2013).

3.10.2 Determining accuracy

There are many ways how to determine the accuracy of the predictive model. We chose to compare the calculated outcomes and the true outcomes in three different ways explained below. The theory that was used for this section is based on (Nataša, 2016) Our dependent variable is coded as 0 for the non-poor and 1 for the poor. When we compare the predicted values obtained from our models and the true values of poverty we can get four possible scenarios.

Table (3.1): Cross Tabulation of True and Calculated Outcomes

	True Poor(1)	True Non-poor(0)
Predicted Poor(1)	True positives	False positives
Predicted Non-poor(0)	False negatives	True negatives

Source: Natasa ,Plulikova 2016

1. True positives (TP) - these are all the cases where our model predicted the household is poor and the household indeed was poor.
2. True negatives (TN) - these are all the cases where our model categorized the household as non-poor and indeed it was non-poor.
3. False positives (FP) - these are all the cases where our model categorized the household as poor but it was not.
4. False negatives (FN) - these are all the cases where our model classified household as non-poor but indeed it was poor.

When it comes to accuracy, there are three ways how we can look at this table:

1. **Error rate.** Error rate is simply a ratio of the cases which were identified wrongly. For the purposes of this thesis and to better compare the methods, we calculate its opposite, i.e. ratio of all those cases that were classified correctly. In mathematical terms:
$Errorrate = \frac{1}{n}\sum(y_i \neq y_i') \ Accuracy = \frac{1}{n}\sum(y_i = y_i')$

Where y_i is the observed value and y_i' is the predicted value of our poverty indicator.

2. **Sensitivity.** Sensitivity is the probability that we will classify the poor among those that are truly poor. In mathematical terms, sensitivity can be calculated as follows: *Sensitivity* = $\frac{TP}{(TP+FN)}$···(3.50)

3. **Specificity.** Specificity is the fraction of how many of the non-poor were classified as non-poor. In mathematical terms, specificity can be calculated as follows: *Specificity* = $\frac{TN}{(TN+FP)}$…………………….(3.51)

3.10.3 ROC Curve

A measure of goodness-of-fit often used to evaluate the fit of a logistic regression model is based on the simultaneous measure of sensitivity (True positive) and specificity (True negative) for all possible cutoff points. First, we calculate sensitivity and specificity pairs for each possible cutoff point and plot sensitivity on the y axis by (1-specificity) on the x axis. This curve is called the receiver operating characteristic [Receiver Operating Characteristic (ROC)] curve. The area under the ROC curve ranges from 0.5 and 1.0 with larger values indicative of better fit (Kotze, & Zeeman, 2014).

Logistic regression is a method for fitting a regression curve, *y* = *f(x)*, when *y* consists of binary coded (0, 1- -failure, success) data. When the response is a binary (dichotomous) variable and *x* is numerical, logistic regression fits a logistic curve to the relationship between *x* and *y*. Logistic curve is an S-shaped or sigmoid curve, often used to model population growth (Eberhardt & Breiwick, 2012). A logistic curve starts with slow, linear growth, followed by exponential growth, which then slows again to a stable rate.

The area under the ROC curve (AUC) The ROC curve can be summarized by the area under the curve (AUC), computed by the trapezoidal rule (base times the median altitude):

$$A = \sum_{t=1}^{n} [x_{i+1} - x_i][(y_{i+1} + y_i)/2)]$$ …………………………. (3.52)

Where, the I are the thresholds where the curve is computed, note that the area under the diagonal is 0.5, so the ROC curve must define an area at least that large. The ROC area then measures the discriminating power of the model:

the success of the model incorrectly classifying sites that did and did not actually change (Rossiter, & Loza, 2008).

3.11 Residual Diagnostics

So far, we have discussed some summary statistics to measure the goodness of fit for the regression model. That is a single number is used to summarize considerable information. Before concluding that the model "fits", it is a common practice to examine other measures to see if fit is supported over the entire set of covariate space, i.e. to examine the influences of individual observations (Wu, & Zhang, 2006).

Residuals are the discrepancies between the data values and their predicted values from the fitted model. A residual analysis detects outliers, identifies influential observations, and diagnoses the appropriateness of the logistic model. An analysis of the residuals should be conducted before a regression model is used.

Unfortunately, the residuals are more difficult to define in logistic regression than in regular multiple regression because of the nonlinearity of the logistic model and because more than one regression equation is used. The discussion that follows provides an introduction to the residuals that are produced by the logistic regression procedure. (Pregibon, 1981) presented this material for the case of the two-outcome logistic regression. Extensions of Pregibon's results to the multiple-group case are provided in an article by Lesaffre and Albert (1989) and in the book by Hosmer and Lemeshow (1989). Lesaffre and Albert provide formulas for these extensions. On the other hand, Hosmer and Lemeshow recommend that individual logistic regressions be run in which the each group is treated separately. Hence, if you have three outcomes A, B, and C, you would run outcome A versus outcomes B and C, outcome B versus outcomes A and C, and outcome C versus outcomes and A and B. You would conduct a residual analysis for each of these regressions using Pregibon's two-outcome formulas.

3.11.1 Pearson Residuals

Identify observations that are not well explained by the model, Pearson residuals are components of the Pearson Chi-square statistic (Wu, and Zhang, 2006).

One popular alternative to the simple residuals are the Pearson residuals which are so named because they give the contribution of each observation to the Pearson chi-square goodness of fit statistic. When the values of the independent variables of each observation are unique, the formula this residual is:

$$\chi'_j = \pm\sqrt{\sum_{g=1}^{G}\frac{\left(y_{gj}-p_{gj}\right)^2}{p_{gj}}}, \quad j=1,2,\cdots,N \qquad \text{……………………….(3.53)}$$

The negative sign is used when $y_{gj} = 0$ and the positive sign is used when $y_{gj} = 1$.When some of the observations are duplicates and the database has been collapsed the formula is:

$$\chi_j = \pm\sqrt{\sum_{g=1}^{G}\frac{\left(w_{gj}-n_j p_{gj}\right)^2}{n_j p_{gj}}}, \quad j=1,2,\cdots,J$$

Where the plus (minus) is used if $w_{gj}/\ n_j$ is greater (less) thanp_{gj}. By definition, the sum of the squared Pearson residuals is the Pearson chi-square goodness of fit statistics. That is,

$$\chi^2 = \sum_{j=1}^{J}\chi_j^2$$

3.11.2 Student zed Pearson Residuals

$$r_{SP_i} = \frac{r_{P_i}}{\sqrt{1-h_{ii}}}, \text{ where}$$

h_{ii} is the ith diagonal element of the matrix $\mathrm{H} = \hat{\mathrm{W}}^{1/2}X(X'\hat{W}X)^{-1}X'\hat{W}^{1/2}$,

$\hat{W}$ is the $n \times n$ diagonal matrix with elements $\hat{\pi}_i(1-\hat{\pi}_i)$

3.11.3 Deviance Residuals

Components of the deviance Chi-square, which is another goodness-of-fit statistic based on the log likelihood function Remember that the deviance is -2 times the difference between log likelihoods of a reduced model and the saturated model. The deviance is calculated using:

$$
\begin{aligned}
D &= -2\left[L_{\text{Reduced}} - L_{\text{Saturated}}\right] \\
&= -2\left[\sum_{j=1}^{N}\sum_{g=1}^{G} y_{gj}\ln\left(p_{gj}\right) - \sum_{j=1}^{N}\sum_{g=1}^{G} y_{gj}\ln\left(y_{gj}\right)\right] \\
&= -2\left[\sum_{j=1}^{N}\sum_{g=1}^{G} y_{gj}\ln\left(p_{gj}\right)\right] \\
&= \sum_{j=1}^{N}\left[2\sum_{g=1}^{G} y_{gj}\ln\left(\frac{1}{p_{gj}}\right)\right]
\end{aligned}
$$

……… ………… (3.54)

This formula uses the fact that the saturated model reproduces the original data exactly and that, in these sums, the value of 0 ln(0) is defined as 0 and that the ln(1) is also 0.
The deviance residuals are the square roots of the contribution of each observation to the overall deviance. Thus, the formula for the deviance residual is

$$d'_j = \pm\sqrt{2\sum_{g=1}^{G} y_{gj}\ln\left(\frac{1}{p_{gj}}\right)}, j = 1,2,\cdots, N$$

…………………………(3.55)

The negative sign is used when $y_{gj} = 0$ and the positive sign is used when $y_{gj}=1$.

When some of the observations are duplicates and the database has been collapsed (see Data Configuration above) the formula is

$$d_j = \pm\sqrt{2\sum_{g=1}^{G} w_{gj}\ln\left(\frac{w_{gj}}{n_j p_{gj}}\right)}, \quad j = 1,2,\cdots, J$$

………………………………….(3.56)

Where the plus (minus) is used if wREF(g), j / n j is greater (less) than pREF (g), j . Note that this is the formula used by **NCSS**. By definition, the sum of the squared deviance residuals is the deviance. That is,

$$D = \sum_{j=1}^{J} d_j^2$$

3.12 Predicted Probabilities

This section describes how to calculate the predicted probabilities of outcome-group membership and associated confidence intervals.

Recall that the regression equation is linear when expressed in logit form. That is,

$$\ln\left(\frac{p_g}{p_1}\right) = \ln\left(\frac{P_g}{P_1}\right) + \beta_{g1}X_1 + \beta_{g2}X_2 + \cdots + \beta_{gp}X_p$$
$$= \ln\left(\frac{P_g}{P_1}\right) + \mathrm{XB}_g \quad \text{…………………………..(3.57)}$$

The adjustment for the prior probabilities changes the value of the intercepts, so this expression may be simplified to

$$\ln\left(\frac{p_g}{p_1}\right) = \beta_{g1}X_1 + \beta_{g2}X_2 + \cdots + \beta_{gp}X_p$$
$$= \mathrm{XB}_g$$

If we assume that the intercepts have been appropriately adjusted. Assuming that the estimated matrix of regression coefficients is distributed asymptotically as a multivariate normal, the point estimates of this quantity for a specific set of X values is given by

$$l_j = \ln\left(\frac{p_g}{p_1} \mid \mathrm{X}_j\right) = \mathrm{X}_j\hat{B}_g \quad \text{……………………………………….(3.58)}$$

And the corresponding confidence interval is given by

$$l_j \pm z_{\alpha/2}\left(\mathrm{X}_j'\mathrm{V}_g\mathrm{X}_j\right)$$

Where

V_g is that portion of the covariance matrix V(Bˆ) that deals with the g^{th} regression equation. When there are only two groups, these confidence limits can be inverted to give confidence limits on the predicted probabilities as

$$\frac{1}{1+e^{\mathrm{X_j\hat{B}} \pm z_{\alpha/2}\sigma_B}} \text{ and } \frac{e^{\mathrm{X_j\hat{B}} \pm z_{\alpha/2}\sigma_B}}{1+e^{\mathrm{X_j\hat{B}} \pm z_{\alpha/2}\sigma_B}} \quad \text{………………………….. (3.59)}$$

Where, $\sigma_{\mathbf{B}} = X'_j VgX_j$

When there are more than two groups, the confidence limits on the logits are still given by$l_j \pm z\alpha / 2(X' jVgX j)$

However, this set of confidence limits of the logits cannot be inverted to give confidence limits for the predicted probabilities.

We have found no presentation that gives an appropriate set of confidence limits. In order to provide an approximate answer, we provide approximate confidence limits by applying the inversion as if there were only two groups. This method ignores the correlation between the coefficients of the individual equations. However, we hope that it provides a useful approximation to the confidence intervals.

CHAPTER FOUR

ANALYSIS OF THE DATA AND INTERPRETATION

4.1Preface

4.2 Background to Central Bureau of Statistics

4.3 Data Source

4.4 Sample Design for NHBPS

4.5 Statistic Methods

4.6 Variables of Study

4.7 Data Analysis, Result and Discussion

ANALYSIS OF THE DATA AND INTERPRETATION

4.1Preface

This chapter presents the background to CBS, statistic methods and variables that we will be utilized in this study. This chapter will be depends entirely on quantitative and qualitative raw data to analysis.

4.2 Background to Central Bureau of Statistics

The Central Bureau of Statistics is the official in charge of Sudan for the collection of statistical data and information, preparation and processing, dissemination and giving nature official figures Statistical It is also in charge of the implementation of statistical and data collection of various kinds, specialties and levels and performs a lot of (general censuses and statistical surveys) and has special versions of the results of these censuses and surveys The statistical Yearbook issued the end of each year, which includes the latest statistics and indicators.

4.3 Data Source

This research depends entirely on micro-level and household's data that will be collected from National Household Budget and Poverty Survey undertook by Central Bureau of Statistics (2015), which is the most recent, available at the time, this study. A sample size of 13800 was drawn representatively across all eighteen States of Sudan, 11953 households were surveyed during the three rounds of data collection with response rate 87%. But use her only 50% were randomly selected for researchers about 5965 household.

4.4 Sample Design for NHBPS

The sample design for the National Household Budget and Poverty Survey 2015 (NHBPS) was a stratified three-stage cluster sample. The sampling frame for Sudan was the preliminary count of households by enumeration area (EA) of the Sudan Fifth Population and Housing Census 2008. The sample size was determined for obtaining reliable estimates for key survey indicators at the state level and for the urban and rural domains at the national level.

A sample of 690 EAs (clusters) was selected at the third sampling stage for each of the 18 states of Sudan. Therefore the estimated total sample size was 13800 households for Sudan. The survey aims to provide information on various socio-economic aspects such as household and housing characteristics, employment pattern, education level and income distribution etc. of the rural as well as urban market centers where the households were situated.

4.5 Statistic Methods

The research analyses the NHBPS data using the SPSS and STATA statistical Packages and analysis will be carried out in two stages that will attain the aims and answers the hypothesis. The first stage descriptive statistics such as frequencies, percentages, and cross-tabulations will be used to analyze relationships between demography characteristics, socioeconomic characteristics and household’s poverty and, also the significance of the association is determined by the Pearson’s chi-square value. The second stage logistic regression model main consider in the analysis and build the model of the study by using SPSS version 20. This program computes both regular (binary) logistic regression and multiple-group logistic regression on both numeric and categorical variables. The level of significance used in all the statistical tests run is the conventional 5%. In constructing the model, $p < 0.05$ is a reasonable guideline for preliminary inclusion.

4.6 Variables of Study

In this study, the dependent variable should be dichotomous. Independent variables can be interval level or categorical; if categorical, they should be indicator coded. Binary logistic regressions support only a single dependent variable, and the response variable can have only two categories, but never a continuous variable. For logistic regression model used in this study, household expenditures per capita were measured as average household income in SDG per year are considered dependent variable. This is calculated considering both food and non-food expenditure including in-kind values in the household. These were codified in the poor (1) and non-poor (0).

Families with yearly per capita consumption expenditure less than the poverty line are considered poor and those with costs greater than the poverty threshold are considered non-poor. The set of independent variables, that are included in the model of the determinants of poverty in Sudan some important demography, social and economic factors. These will be broken down into smaller specific variables that can easily be measured and understood they include such; place of residence, household size, sex of household head, age of household head, dependence ratio, Can read and write with understanding, level education of household head, vocational training / craft, marital status, disability type of head, Worked at least one hour for profit in cash or kind last 10 days, dwelling Type , main tenure status for the dwelling and Main source of livelihood of households are considered as covariates in the binary logistic regression model. The data description of the fourteen independent variables and the dependent variable are provided in Table (4.1). The operational definitions of the variables included in the model are defined as follows table (4.2):

Table (4.1) Characteristic of the Dependent Variable

Dependent Variable	Characteristic
Poverty	$\begin{Bmatrix} (1 = poor)\ , if\ the\ household\ is\ below\ the\ poverty\ line\ (6082) \\ (0 = non\ poor), if\ \ otherwise \end{Bmatrix}$

Source: The researcher own table2018

Table (4.2) Definitions of the Independent Variables

N	Demography Variable	Abbreviations	Definition	Characteristic
1	Place of residence	PR	whether a household is located in the rural or urban area	2 categories are represented by one binary variable as follows: Dummy, Rural = 1, Urban = 0
2	household size	HS	Total household members	Continuous
3	sex of household head	SHH	Sex of household head (male or female).	Dummy, Female = 1 , Male =0

4	age of household head	AHH	Age of household head (years)	Continuous
5	dependence ratio of household	DR	DR of number of members (<15 years and >64 years) to household size and treated as a continuous variable.	Continuous
	Social Variables			
6	Can read and write with understanding	CRW	Can read and write with understanding any sentences	Yes = 0, No = 1
7	Highest level of school ever completed	HLS	refers to the highest level of schooling that a person has reached	Category: 1- no qualification 2- primary 3- intermediate 4- secondary 5- university 6- postgraduate 7-khalowa
8	Vocational training /craft Of household head	VT	Training that emphasizes the knowledge and skills needed for a specific trade, craft or job function.	Yes =0, No = 1
9	marital status of household head	MS	The term describes whether the head of household is married or not.	Category: 1- never married 2- married 3- Widow 4- divorce
10	Disability type of household head	Dis	Suffering from any type of disability that prevents from doing usual work.	Yes = 1, No = 0

	Economic Variables			
11	Worked at least one hour for profit	Work	It means work at least one hour for profit	Yes = 1, No = 0
12	Dwelling Type	DT	A durable item that can be used for more than one year.	Category: 1-gottiya or tent 2-apartment 3-house of one floor 4-multi-story floor 5-incomplete
13	Main tenure status for the dwelling	MTSD	Home ownership	Category: 1-owned 2-rented 3-housing provided as part work 4-free
14	Households main source of livelihood	MSL	Refers to their "means of securing the basic necessities –food and non food	Category: 1-crop farming 2-animal husbandry 3-wages and salaries 4-owned income 5-remittances and aid 6-transfers from members

Source: The researcher own table, 2018

4.7 Data Analysis, Results and Discussion

In this chapter, the findings of the study are using descriptive statistics and Logistic regression to analysis and estimates of factors affecting poverty by using SPSS statistical Packages (version 2o) and STATA.

4.7.1 Descriptive Analysis

The descriptive statistics which include mean, standard deviation, minimize and maximize of the various demography, social and economic variables analyzed in the study, and applied each one by cross tabulation analysis.

4.7.1.1 Demography Factors

This section presents demography status provides both the result cross-tabulated and logistic regression analysis by characteristics of the household, like a place of residence, household size, sex of household head, the age of household head and a dependency ratio of the household are described below.

Table (4.3) Descriptive Statistics of Demography Variables and Poverty

Poverty household status			place of residence	Household size	Sex of household head	Age of household Head	Dependency ratio of household
Non poor	N	Valid	4060	4060	4060	4060	3990
		Missing	0	0	0	0	70
	Mean		-	5.12	-	46.81	99.500
	Std. deviation		-	2.233	-	15.194	94.0600
	Minimize		-	1	-	15	.0
	Maximize		-	17	-	95	1100.0
Poor	N	Valid	1904	1904	1904	1904	1894
		Missing	1	1	1	1	11
	Mean		-	7.37	-	46.04	163.047
	Std. deviation		-	2.322	-	13.155	120.5216
	Minimize		0	2	0	17	0
	Maximize		1	20	1	95	1000

Source: Prepared by researcher from, 2018

The survey included a total of 5965 households. Descriptive statistics on demography characteristics of the household affecting poverty levels in Sudan are discussed. In the above table (4.3) shows that some cases are missing observations, these from a non-response to the survey question، also in case of result of calculation operations of dependency ratio give missing value. The largest number of missing values 70 households is in non poor response to the dependency ratio variable, but only 11 household from poor responses. The average of household size of poor was found to be 7.37 and non poor households were 5.12person.

While the minimum and the maximum number of members in the family for poor households was found to be 2, 20 members and for non poor was 1, 17 size respectively. The average age of household heads for poor and non-poor households was 46.04 and 46.81 years respectively, with a standard deviation of 13.16 and 15.19, and the minimum and the maximum age for poor and non poor household heads stood at 17, 95 years and 15, 95 years respectively. The average dependency ratio of households for poor and non poor was 163.05% and 99.5% percent respectively. The minimum and maximum dependency ratio for poor stood at 0% and 1000%, similarly for non poor at 0% and 1100% respectively.

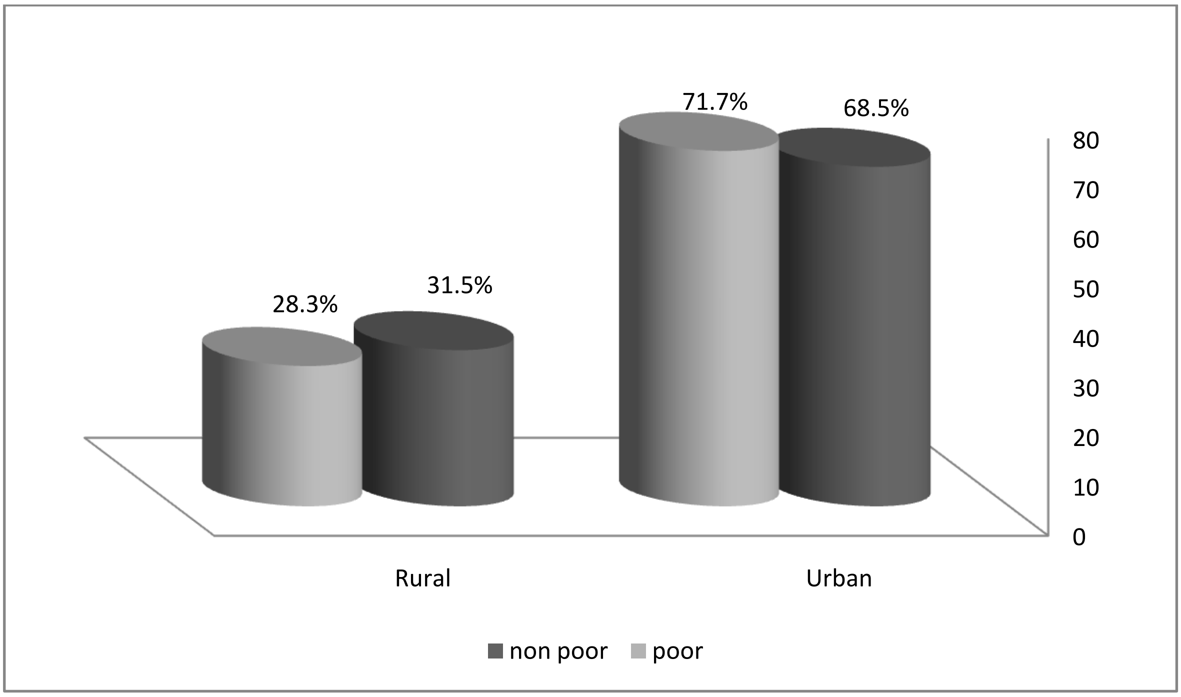
Table (4.4) Association between poverty and place of residence

			Place of residence (PR)		Total
			Urban	**Rural**	
Poverty household status	Non poor	N	1278	2782	4060
		Percent % Within poverty	31.5%	68.5%	100.0%
	poor	N	539	1366	1905
		Percent % Within poverty	28.3%	71.7%	100.0%
Total		N	1817	4148	5965
		Percent % Within poverty	30.5%	69.5%	100.0%

Source: Prepared by the researcher by using SPSS package, 2018

According to the association between poverty and type of place of residence in the above table (4.4), the result shows that most cases of the very poor are in rural areas about 71.7% and 28.3% in the urban areas. See figure (4.1) below explain the relationship between poverty and place of resident.

Figure (4.1) Association between poverty and place of resident



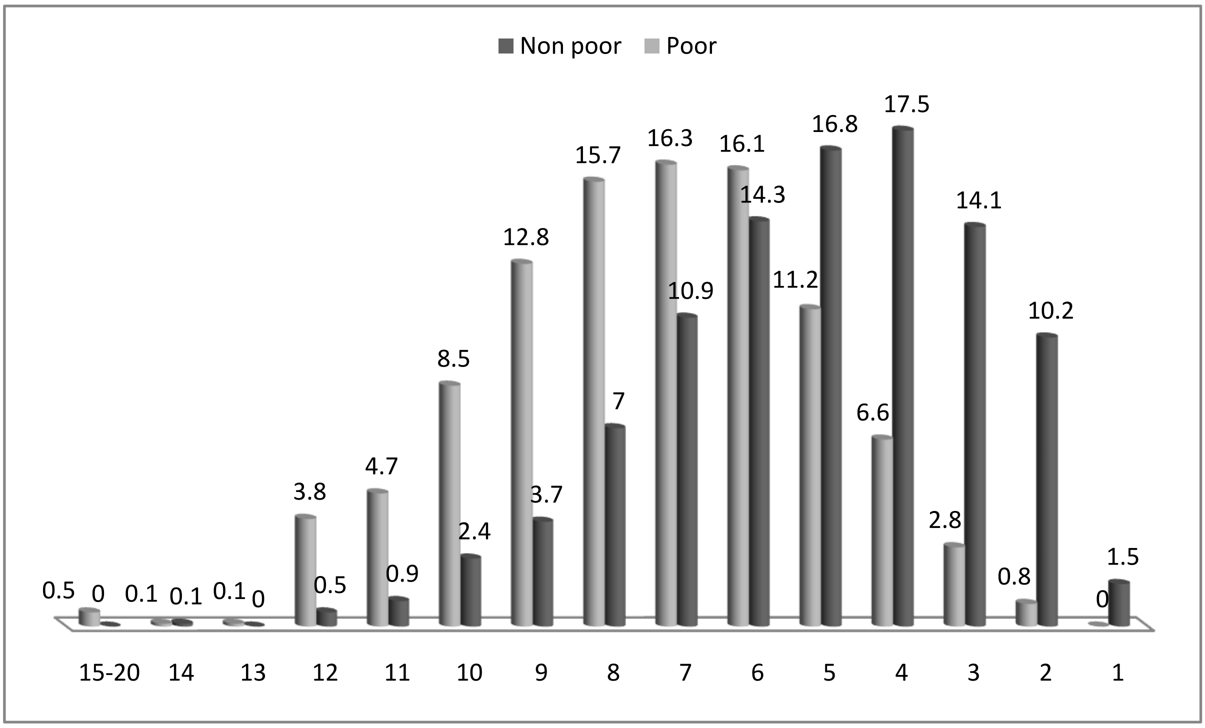
Source: Prepared by researcher by using Excel, 2018

Table (4.5) Relationship between poverty and household size

			Poverty household status		Total
			Non poor	Poor	
HH Size	1	Count % within poverty	60 1.5%	0 0.0%	60 1.0%
	2	Count % within poverty	413 10.2%	16 0.8%	429 7.2%
	3	Count % within poverty	571 14.1%	54 2.8%	625 10.5%
	4	Count % within poverty	712 17.5%	125 6.6%	837 14.0%
	5	Count % within poverty	683 16.8%	213 11.2%	896 15.0%
	6	Count % within poverty	581 14.3%	306 16.1%	887 14.9%
	7	Count % within poverty	444 10.9%	310 16.3%	754 12.6%
	8	Count % within poverty	283 7.0%	300 15.7%	583 9.8%
	9	Count % within poverty	151 3.7%	243 12.8%	394 6.6%
	10	Count % within poverty	99 2.4%	161 8.5%	260 4.4%
	11	Count % within poverty	36 0.9%	89 4.7%	125 2.1%
	12	Count % within poverty	21 0.5%	73 3.8%	94 1.4%
	13	Count % within poverty	2 0.0%	4 0.1%	6 0.1%
	14	Count % within poverty	3 0.1%	1 0.1%	4 0.1%
	15-20	Count % within poverty	1 0.0%	10 0.5%	11 0.2%
	Total		4060 100%	1905 100%	5965 100%

As described in the table (4.6), the Large households tend to be associated with higher poverty. The result shows that poverty is highest for a household with 7 and 6 members and become lowers for households of other sizes. Household sizes tend to be higher for poor households since the size for poor households is 7 compared to 4 for non-poor households. The statistics show that there is an increase in poverty with an increase in household size, this relation so clear in the below figure (4.2).

Figure (4.2) Association between poverty and household size



Source: Prepared by researcher by using Excel, 2018

Table (4.6) Relationship between Poverty and sex of household head

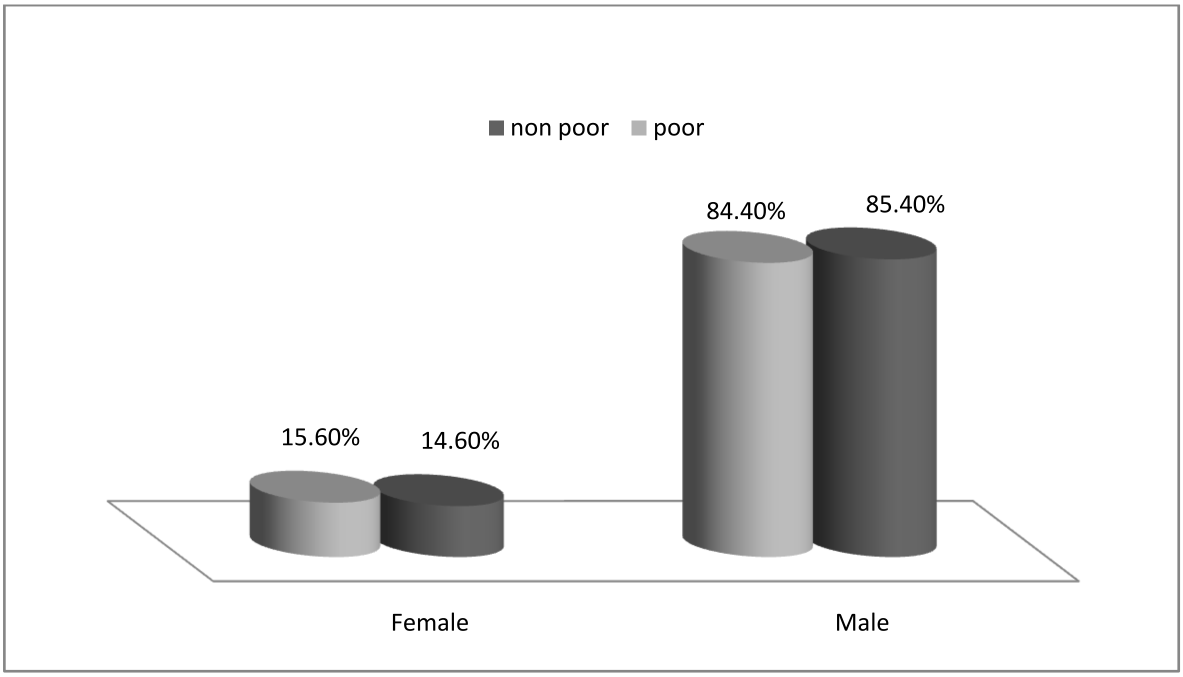
			Sex of household head (SHH)		Total
			Male	Female	
Poverty household status	Non poor	N	3468	592	4060
		Percent % Within poverty	85.4%	14.6%	100.0%
	poor	N	1608	297	1905
		Percent % Within poverty	84.4%	15.6%	100.0%
Total		N	5076	889	5965
		Percent %	85.1%	14.9%	100.0%

Source: Prepared by the researcher by using SPSS package, 2018

In the above table (4.6) the study revealed that majority of 85.1% of the household heads was male. This is usually typical and natural household structure in a traditional African setting and in most other continents of the world. Females only become the household head in the event of death of the husband, or outright divorce. Also the result shows that the association between poverty and sex of households, it was revealed that households were headed by women were more likely to be poor than those headed by their males, and about 84.4% of male and 15.6% of female are poor. Similarly, the proportion of the non-poor female headed households is 14.6% from 592 household while male headed household is 85.3%. More explains in figure (4.3)

s

Figure (4.3) Association between poverty and sex of household head



Source: Prepared by researcher by using Excel, 2018

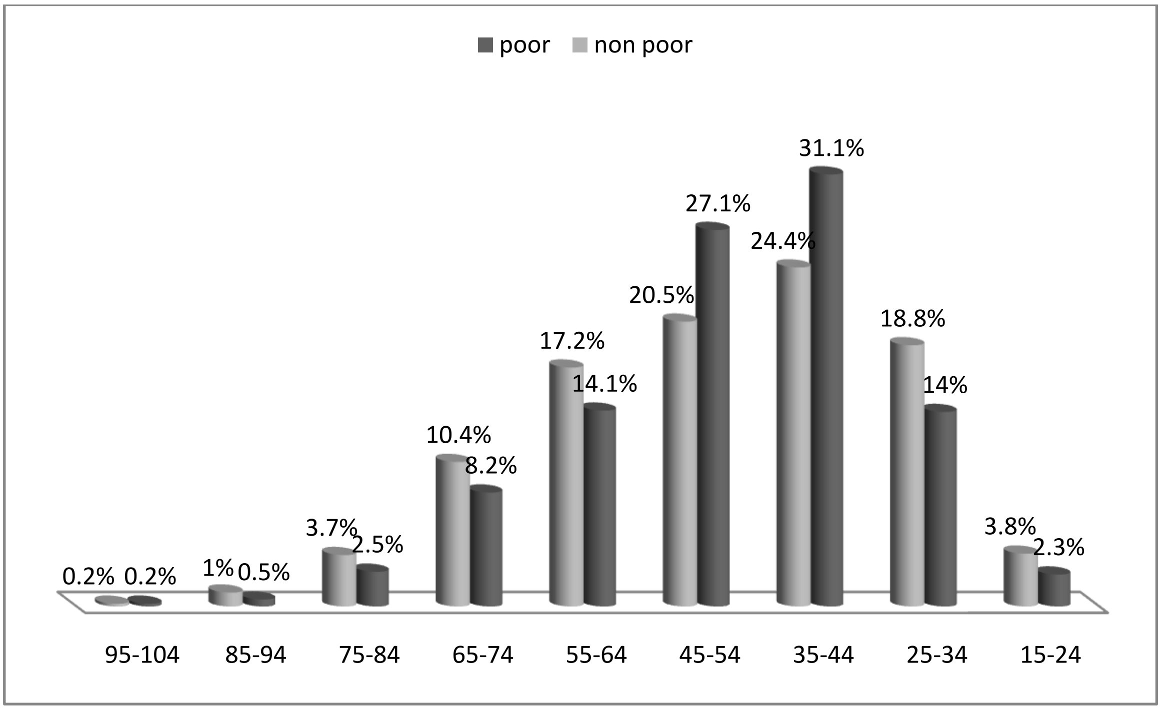
Table (4.7) Relationship between poverty and age of household head

			Poverty household status		Total
			Non poor	Poor	
Age of HH	15-24	N	154	44	198
		% within Poverty	3.8%	2.3%	3.3%
	25-34	N	765	266	1031
		% within Poverty	18.8%	14.0 %	17.3%
	35-44	N	989	593	1582
		% within Poverty	24.4%	31.1 %	26.5%
	45-54	N	834	516	1350
		% within Poverty	20.5%	27.1 %	22.6%
	55-64	N	698	269	967
		% within Poverty	17.2%	14.1 %	16.2%
	65-74	N	421	156	577
		% within Poverty	10.4%	8.2%	9.7%
	75-84	N	150	48	198
		% within Poverty	3.7%	2.5%	3.3%
	85-94	N	42	9	51
		% within Poverty	1.0%	0.5%	0.9%
	95-104	N	7	3	10
		% within Poverty	0.2%	0.2%	0.2%
Total		N	4060	1904	5964
		% within Poverty	68.1%	100.0 %	100.0%

Source: Prepared by the researcher by using SPSS package, 2018

According to association between poverty and age of household head in the above table (4.7), the results show that 31.1% of the more poorest households are aged between 35 and 44 year, as compared to 25.47% of the poor households head who are over 54 year and 16.28% of the poor who are less than 35 years, while 58.25% of the poor's between 35 and 54 years, and more poor household head in age 40 as depicted in figure (4.4) below .Similarly, the results show that 24.4% of the non- poor households are aged between 35 and 44 as compared to 32.56% of the non-poor households who are over 54years and 22.64% of the non poor who are less than 35 year and more non poor household head in age 35. The results show that poverty increases with young-aged of the household heads but in 54 years and above indicate that the age of the heads increases poverty levels reduce, or poverty decreases with the increasing age of the household head and then increases again at old age in Sudan.

Figure (4.4) Relationship between age of household head and poverty



Source: Prepared by researcher by using Excel, 2018

Table (4.8) Relationship between poverty and Dependency Ratio (DR)

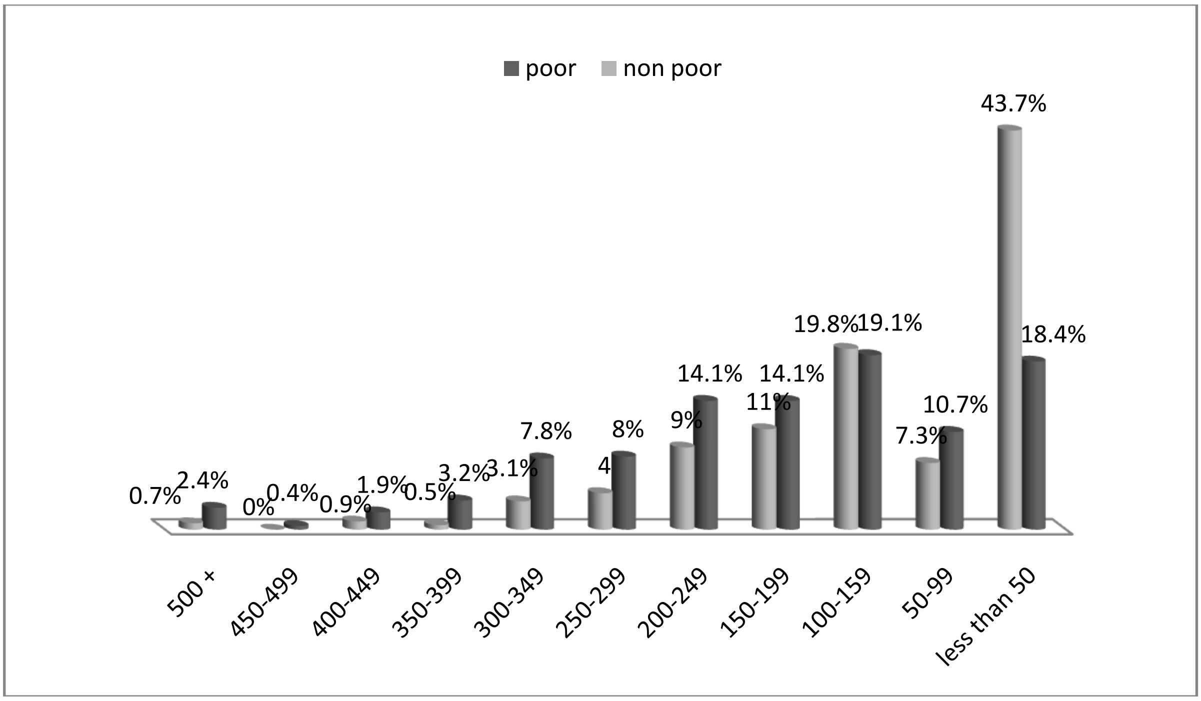
			Poverty household status		Total
			non poor	poor	
DR	less than 50	N	1743	348	2091
		% within poverty	43.7%	18.4%	35.5%
	50-99	N	291	202	493
		% within poverty	7.3%	10.7%	8.4%
	100-149	N	792	361	1153
		% within poverty	19.8%	19.1%	19.6%
	150-199	N	438	268	706
		% within poverty	11.0%	14.1%	12.0%
	200-249	N	360	267	627
		% within poverty	9.0%	14.1%	10.7%
	250-299	N	159	152	311
		% within poverty	4.0%	8.0%	5.3%
	300-349	N	125	147	272
		% within poverty	3.1%	7.8%	4.6%
	350-399	N	21	60	81
		% within poverty	0.5%	3.2%	1.4%
	400-449	N	35	36	71
		% within poverty	0.9%	1.9%	1.2%
	450-499	N	0	8	8
		% within poverty	0.0%	0.4%	0.1%
	more than 500	N	26	45	71
		% within poverty	0.7%	2.4%	1.2%
Total		N	3990	1894	5884
		% within poverty	100.0 %	100.0 %	100.0 %

Source: Prepared by the researcher by using SPSS package, 2018

From the table (4.8) above the dependency ratio shows that the households with a high dependency ratio are poorer than the household with a low dependency ratio.

As described that 19.1% of the poorest households are grouped between 100 and 149% for dependency ratio. See figure (4.5) blow describe clearly to dependency ratio and also shows that the average of dependency ratio 119.96% with stander deviation 107.49. A larger households with a high portion of dependents are more prone to poverty is plausible.

Figure (4.5) Relationship between Dependency Ratio (DR) and Poverty



Source: Prepared by the researcher by using Excel, 2018

1- Estimation binary logistic regression for demography factors

The logistic regression technique has been applied to evaluate the demography characteristics of the household's head and household characteristics as the determinants of household poverty in Sudan. The demography factors include place of residence, household size, sex of household head, the age of the Household head and dependency ratio.

Table (4.9) Crosstabs of Pearson chi-square statistics test for association between demographic characteristics and Poverty

Variable	Pearson Chi-Square	Likelihood Ratio	df	Sig
PR	6.352	6.401	1	.012
HS	1082.681	1171.799	17	.000
SHH	1.041	1.o35	1	.311
AHH	97.531	98.634	8	.000
DR	515.114	527.017	10	.000

Source: Prepared by researcher by using SPSS package 2018

At 5% level of significance, the following variables are statistically insignificant: The chi-square values of the remaining variables make the null hypothesis to be rejected and the alternative hypothesis to be adopted. The inference is that from table (4.9), we observe that there is a very strong association between place of residence, household size, age of household head, and dependency ratio with Poverty household status, except sex of household head was not significant at 5% level of significance.

Table (4.10) Coefficients and Wald test for logistic regression on the poverty and demography data

	B	S.E	Wald	df	Sig	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step1[a] PR(1)	-.065	.070	.864	1	.353	.937	.817	1.075
HS	.410	.015	718.322	1	.000	1.507	1.463	1.553
SHH(1)	-.566	.093	37.402	1	.000	.568	.473	.681
AHH	-.049	.024	4.303	1	.038	.952	.909	.997
DR	.160	.015	113.782	1	.000	1.173	1.139	1.208
Constant	-3.162	.150	444.194	1	.000	.042		

Source: Prepared by the researcher by using SPSS package, 2018

Variable(s) entered on step 1: PR, HS, SHH, AHH, and DR. a

Table (4.10) shows that out of the fifth identified variables only one variable was not significant in explaining whether household status is poor or not poor. As can be seen in the above table, the variables HS and DR have odds ratios greater than one, which means that these variables are positively correlated with the probability of being poor, while those variables SHH and AHH which have odds ratios lower than one are inversely correlated with the probability of being poor.

The confidence interval for the odds ratio of PR includes the number one, which means that this variable has not statistically significant effect on the probability of poverty.

If we look at the table (4.10) we see that household size (HS) is a statistically significant variable (B = 0.410, Wald = 718.322, P =.000) and its positive coefficient indicates that with increasing household size increases the probability that the household be poor. Thus, even though the coefficient (B= -.566, Wald=37.402, P-value=.000) for the Sex of household head (SHH) variable is negative and statistical significant. The gender of head variable is an important factor in explaining the poverty status of the family but the negative coefficient indicates that households headed by female have lower probability of being poor than male-headed households.

The coefficient for the variable age of household head (AHH) is negative and statistically significant variable (B= -.o49, Wald= 4.303. P=.000). This means that there is an established negative relationship between age of household head and the per capita expenditure of the household. The age of household heads grows older; the per-capita expenditure/income of the household reduces, thus, increasing the level of poverty in the household. Thus, we can see that an increase of one year in the age of the head decreases the odds of being poor by almost 95.2%. We found that there is a strong and statistically significant inverse relationship between poverty and age of the head.

The coefficient of the dependent ratio (DR) is positive and statistically significant (B=.160, Wald= 113.782, P= .000). The odds ratio of the variable dependency ratio shows a contribution of 20.8% in increasing the likelihood of being poor whereas household size (HS) contributes 55.3%. Therefore the majority of households fell in poverty because of having large families with many dependants being children or elderly at unproductive age.

Type of place of residence (PR) had an insignificant impact on poverty. It had been found that (B = -.o65, Wald = .864, p = .253). This means that there is a negative relationship between the type of place of residence and poverty. These finding confirmed the conclusion of other studies, such as (Garza, 2015).

The inference that the households who are residence in the urban areas and rural areas influences by the poverty according to our income and levels of per capita expenditures.

Table (4.11) Omnibus Tests of Model Coefficients

		chi-square	Df	sig
Step 1	Step	1305.466	5	.000
	Block	1305.466	5	.000
	Model	1305.466	5	.000

Source: Prepared by the researcher by using SPSS package, 2018

In the above table (4.11) we have added all five explanatory variables in one block and therefore have only one step. This means that the chi-square values are the same for step, block, and model. Here the chi-square is highly significant (chi-square=1305.466, df =5, p-value <.000) so our model is significantly better, which indicates the accuracy of the model improves when we add our explanatory variables.

Table (4.12): Summary measures of goodness-of-fit statistics of the model with selected covariates

Summary Statistic test	**Value**	**df**	**P-value**
Hosmer - Lemeshow	32.654	8	.000
LR chi-square	1305.47	8	.000
Log likelihood	3044.101		
Cox and Snell R^2	0.199		
Nagelkerke R^2	o.278		
Pseudo R^2	0.1766		

Source: Prepared by the researcher by using SPSS package and STATA, 2018

The goodness-of-fit measures how effectively the model describes the response variable. According to the table (4.12), the Hosmer-Lemeshow (H-L) test that yields a $\chi^2$of 32.654 and was significant Suggesting that there was lack of fit the model of the data. Thus we reject the null hypothesis that the model not represented the data well.

The log likelihood yields a χ^2 of 3044.101and significant at (p> 0.05) which also give a good fit for the model to the data and thus the null hypothesis was also tenable for the model.

Values of statistics Cox Snell R^2 and Nagelkerke R^2 parameters are 0.199 and 0.278 which indicate that the model explains 19.9% to 27.8% of the variance in the outcome. The model has a pseudo R^2 of 0.177 which means that 17.7% of the variation in the dependent variable is due to the variations in the independent variables.

Table (4.13) Correct Classification Table [a] of the Model

Observed		Predicted		
		Poverty household status		Percentage Correct
		Non Poor	Poor	
Poverty household status	Non poor	3541	449	88.7
	poor	1101	793	41.9
Overall Percentage				73.7

Source: Prepared by the researcher by using SPSS package, 2018

The cut value is .500[a]

The classification table (4.13) shows that the model makes a correct prediction 73.7% of the cases compared to 67.8% in the null model, a marked improvement. Of the 3990 households with non-poor, the model correctly identified 449 households predicted as non-poor by the model are in fact poor. Similarly, of the 1894 households that did have a poor, the model correctly identifies 793 households predicted as poor by the model are in fact non-poor. 41.9% is known as the **sensitivity** of prediction. 88.7% is also known as the **specificity** of prediction.

The logistic model was fitted to the data to test the relationship between the likelihood of a household being poor or non-poor, it's carried out by the enter method, and the result showed that, in the optimal model i:

$$\text{Log}\left(\frac{P}{1-P}\right) = Y = -3.162 - 0.65\ PR\ (1) + 0.410\ HS - 0.566\ SHH\ (1) - 0.049\ AHH + 0.160\ DR \quad \ldots\ldots(4.1)$$

The estimates of the logistic regression are shown in the above Tables. In general, the logit model fitted the data quite well.

The chi-square test strongly rejects the hypothesis of no explanatory power and the model correctly predicted 73.7% of the observations. Furthermore, household size, sex of household head, the age of household head and dependency ratio are statistically significant and the sign on the parameter estimate support expectations, except the place of residence, is not significant.

4.7.1.2 Social Factors

This section provides both the result of cross tabulation and logistic regression analysis. Initially a set of predictors such as can read and write with understanding, highest level of school ever completed, ever attended vocational training, marital status and Suffering from any type of disability that prevents from doing usual work.

Table (4.14) Association between Poverty and Can Read and Write with Understanding

			Can read and write with understanding (CRW)			Total
			No, Cannot read and write	Yes, Can read and write	Not stated	
Poverty household status	Non poor	N	1400	2658	2	4060
		% within poverty	34.5%	65.5%	0.0%	100.0%
	Poor	N	861	1038	6	1905
		% within poverty	45.2%	54.5%	0.3%	100.0%
Total		N	2261	3696	8	5965
		%	37.9%	62.0%	0.1%	100.0%

Source: Prepared by the researcher using SPSS package, 2018

As described in the table (4.14) above, about 45.2% of the poorest household heads were illiterate and 54.5% were found being able to read and write from total responses in the sample. Similarly, about 34.5% of non poor heads were illiterate and 65.5% were found being able to read and write.

Table (4.15) Association between Poverty and Highest Level of School ever Completed (HLS)

			Poverty household status		Total
			Non poor	poor	
The Highest level of school ever completed	no qualification	N	528	274	802
		% within poverty	13.0%	14.4 %	13.4%
	Primary	N	622	239	861
		% within poverty	15.3%	12.5 %	14.4%
	Intermediate	N	145	56	201
		% within poverty	3.6%	2.9%	3.4%
	Secondary	N	631	155	786
		% within poverty	15.5%	8.1%	13.2%
	University	N	255	22	277
		% within poverty	6.3%	1.2%	4.6%
	Post graduate	N	25	3	28
		% within poverty	0.6%	0.2%	0.5%
	Khlowa	N	345	227	572
		% within poverty	8.5%	11.9 %	9.6%
	Not stated	N	1509	929	2437
		% within poverty	37.2%	48.8 %	40.9%
Total		N	4060	1905	5964
		% within poverty	100%	100%	100.%

Source: Prepared by the researcher by using SPSS package, 2018

From the above table (4.15) the level of education shows that the poor of the Household heads who had no qualification 14.4%, while 12.5% had primary education and 2.9% completed Intermediate school education. Only 0.2% attended Postgraduate. Those attained secondary and university education 8.1% and 1.2% respectively.

Thus, while the poverty rate for households where the head has no qualification from school is 26.29%. For the not stated column there are about 40.9% non-respondents in questionnaire and data entering errs. The distribution of households by education and poverty status results show that the highest poverty cases have no qualification and Khlowa while those with the higher education have lower cases of poverty.

Table (4.16) Association between Poverty and Vocational Training (VT)

			Ever attended vocational training (VT)			Total
			Never attended	attended	Not stated	
Poverty household status	Non poor	N	1261	2502	297	4060
		% within poverty	31.1%	61.6%	7.3%	100.0 %
	poor	N	758	952	195	1905
		% within poverty	39.8%	50.0%	10.2%	100.0 %
Total		N	2019	3454	492	5965
		%	33.8%	57.9%	8.2%	100.0 %

Source: Prepared by the researcher by using SPSS package, 2018

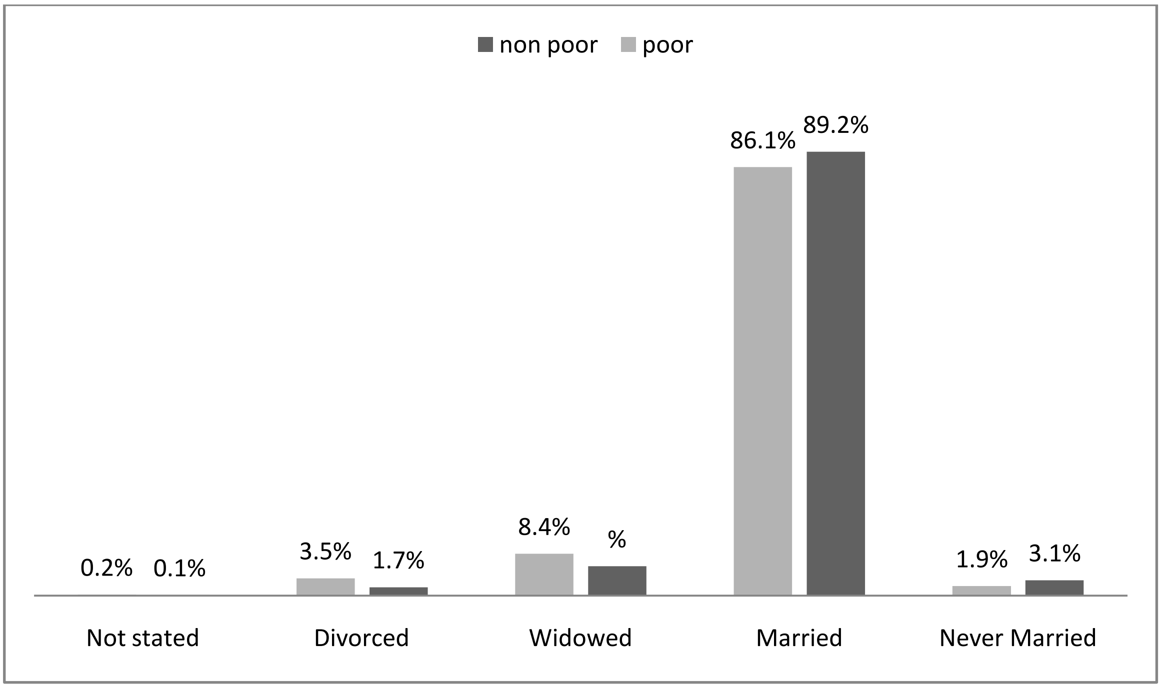
The findings from the table (4.16) showed that about 39.8% represented the poor of the household heads never attended vocational training in Sudan and 50% of the poor attended vocational training. It means that the training is not affected to respondents. Similarly about 61.6% of the non poor of attended vocational training and about 8.2% not respondent to this question.

Table (4.17) Association between poverty and marital status (MS)

			Marital status (MS)					Total
			Never Married	Married	Widowed	Divorced	Not stated	
Poverty	Non poor	N	125	3622	238	71	4	4060
		% within poverty	3.1%	89.2%	5.9%	1.7%	0.1%	100%
	poor	N	37	1639	159	66	3	1904
		% within poverty	1.9%	86.1%	8.4%	3.5%	0.2%	100%
Total		N	162	5261	397	137	7	5964
		% within poverty	2.7%	88.2%	6.7%	2.3%	0.1%	100%n

In this study as shown in the table (4.17) above the data shows that 86.1% of the poor of the household heads are married, followed by 8.4% who are widowed; this may be due to civil wars or diseases in Sudan, 3.5% divorced and only 1.9% never married. While 89.2% of the non poor of the household are married, 5.9 % was widowed and the remaining never married and divorced. It is clear from the result that majority of the poor of the household heads have a high of poverty were married. Not stated numbers indicate that most respondents were unwilling to respond to this question. To support that result see the figure (4.6) below the relation more clearly for poor and marital status.

Figure (4.6) Relationship between marital status (MS) and Poverty



Source: Prepared by researcher, by using Exsal, 2018

Table (4.18) Relationship between Poverty and Suffering any Disability from Work

			Suffering from any type of disability that prevents from doing usual work			Total
			Yes	No	Not stated	
Poverty household status	Non poor	N	241	3818	1	4060
		% within poverty	5.9%	94.0%	0.0%	100.0%
	poor	N	89	1816	0	1905
		% within poverty	4.7%	95.3%	0.0%	100.0%
Total		N	330	5634	1	5965
		%	5.5%	94.5%	0.0%	100.0%

Source: Prepared by the researcher by using SPSS package, 2018

From the table (4.18) above the total respondents who poor and reported a' yes 89 household heads in the questionnaire, whomever 4.7% are poor and about 95.3% reported no were poor. The result shows that households having heads with disability have a less effect on poverty because there 5.9% of household heads of disability are not poor but only 4.7% heads are poor. This may be that families receiving assistances or disability did not prevent them from working.

2- Estimation binary logistic regression for social variables

The logistic regression technique has been applied to estimation the social characteristics of the household's head and household characteristics as the determinants of household poverty in Sudan. The social factors include can read and write with understanding, highest level of school ever completed, ever attended vocational training, marital status and Suffering from any type of disability that prevents from doing usual work.

Table (4.19) Crosstabs of Pearson Chi-square Statistics test and Likelihood Ratio for the Association between Social Characteristics and Poverty

Variable	Pearson Chi-Square		Likelihood Ratio		df
	Value χ^2	P-value	Value χ^2	P-value	
CRW	71.320	.000	70.150	.000	2
HLS	201.891	.000	225.771	.000	7
VT	73.015	.000	72.540	.000	2
MS	36.685	.000	35.536	.000	4
Dis	4.442	.108	4.857	.088	2

Source: Prepared by the researcher by using SPSS package, 2018

To test the association of the variables, in this section we apply the Chi-square test. To perform this, we compare all the explanatory variables with response variable, poverty. The results of the tests are shown in the table (4.19), we observe that there are a very strong association between can read and write with understanding (CRW), the highest level of school ever completed (HLE), ever attended vocational training (VT) and marital status (MS) with poverty household status, except for the variable type of disability (Dis) was not significant at 5% level of significance, but it's significant when included all variables in the logistic regression. Also, can read and write with understanding not significant when analysis with the logistic regression.

Table (4.20) Coefficients and Wald tests for Logistic Regression on the Poverty and Social data

							95% C.I. for EXP(B)	
	B	**S.E**	**Wald**	**df**	**Sig**	**Exp(B)**	**Lower**	**Upper**
Step1 CRW			4.197	2	.123			
CRW(1)	-1.521	.821	3.436	1	.064	.219	.044	1.091
CRW(2)	-1.608	.822	3.830	1	.050	.200	.040	1.002
HLS			125.05	7	.000			
No qu(1)	.018	.202	.008	1	.930	1.018	.686	1.511
Prim(2)	-.283	.205	1.903	1	.168	.753	.504	1.126
Inter(3)	-.274	.248	1.220	1	.269	.761	.468	1.236
Sco(4)	-.728	.211	11.852	1	.001	.483	.319	.731
Univ(5)	-1.800	.290	38.448	1	.000	.165	.094	.292
High(6)	-1.415	.639	4.913	1	.027	.243	.069	.849
Khal(7)	.241	.203	1.402	1	.236	1.272	.854	1.895
VT			4.034	2	.133			
VT(1)	-.169	.117	2.079	1	.149	.844	.671	1.063
VT(2)	-.276	.178	2.409	1	.121	.759	.536	1.075
MS			34.709	4	.000			
Never(1)	-.142	.810	.031	1	.861	.868	.178	4.241
Mar(2)	-1.195	.259	21.308	1	.000	.303	.182	.503
Wid(3)	-.770	.179	18.523	1	.000	.463	.326	.657
Div(4)	-.413	.205	4.060	1	.044	.662	.443	.989
Dis			9.259	2	.010			
Dis(1)	-20.69	40192.9	.000	1	1.00	.000	.000	.
Dis(2)	-.394	.129	9.259	1	.002	.674	.523	.869
Constant	1.950	.838	5.410	1	.020	7.029		

Source: Prepared by researcher by using SPSS package, 2018

Variable(s) entered on step 1: , CRW, HLS, VT, MS and Diss. a

According to the summary statistics from the table (4.20) are showing that the odds ratios of all variables are less than one that puts all variables in negative relation with the poverty status, except two variables (no qualification and khalowa).

The overall result of the table above shows that out of the fifth identified variables only two variables were not significant in explaining whether a household's status is poor or not poor. The study shows that the variable can read and write with understanding does not a significantly affect poverty level probably (Wald=4.187, P-value=.123) and negative relation with the probability of being poor, because of lower education have more chances to be poor.

Generally, the results depict that there was a negative relationship between the probability of being a poor and different level of education. It means that higher levels of education reduce the probability of being poor gradually. If we look at the no qualification, primary, intermediate and khalowa say are not statistically significant, All of them can be classified as weak qualification, these because as the primary and intermediate stages have been integrated into one stage since 1990 and there has been a change in the educational ladder, so they have no any impact to poverty. But the secondary, university and higher education remain an important determinant of household welfare and say are statistically significant variable and are negative coefficients indicate that increased education has a significant impact in reducing the probability of being poor, implying that a higher level of education provides greater opportunities for a better job and, subsequently, a higher income. This implies that education is the important factors in reducing the impact of poverty at the household level. These findings confirmed the conclusions of other studies, such as Bigsten et al. (2003); Achia, (2010), Sarwar et al (2012) and Xhafaj, & Nurja, (2014).

The results also indicate that the vocational training (Wald=4.034, P-value=.133) is not significant in explaining the probability of being a poor and negative relation with the probability of being poor, this was due to lack of adequate training in Sudan.

Furthermore, the marital status is a statistically significant variable (Wald =34.709, P-value=.000) and negatively correlated with responsiveness. Moreover, the study shows that the married household heads have a higher chance of being poor as compared to household heads that are not married. More specifically, the results indicate that the married, widowed and divorced head of households, were significantly more likely to be poor than their never married. This may be as a result of having more dependants depending on the household head. We also found that the never married is a statistically insignificant, those who lived together, and enjoy lower welfare with our families and not married yet, for this they have not impact on poverty. Moreover, we found that the Suffering type of disability is not significant (Wald = 9.259, P-value=.010) in explaining the probability of being a poor and negative relation with a probability of being poor.

Table (4.21) Omnibus Tests of Model Coefficients

		chi-square	df	sig
Step 1	Step	278.923	17	.000
	Block	278.923	17	.000
	Model	278.923	17	.000

Source: Prepared by the researcher by using SPSS package, 2018

In the above table (4.21) we have added all five explanatory variables in one block and therefore have only one step. This means that the chi-square values are the same for step, block, and model. Here the chi-square is highly significant (chi-square=278.923, df = 17, p-value <.000) so our model is significantly better, which indicates the accuracy of the model improves when we add our explanatory variables.

Table (4.22): Summary Measures of Goodness-of-fit Statistics of the Model with Selected Covariates

Summary Statistic test	Value	df	P-value
Hosmer – Lemeshow	4.436	7	.728
LR chi-square	84.70	17	.000
Log likelihood	-3692.917		
Cox and Snell R^2	.46		
Nagelkerke R^2	.64		
Pseudo R^2	.0113		

Source: Prepared by researcher by using SPSS package and STATA, 2018

According to the table (4.22), the variables are significant predictors of poverty ($p < 0.05$) the Goodness-of-fit statistics assess the fit of a model against actual values. The inferential goodness-of-fit test is the Hosmer-Lemeshow (H-L) test that yields a χ^2 of 4.436 and was not significant Suggesting that there was the goodness of fit the model of the data. Thus we accept the null hypothesis that household characteristics and perceptions have an influence on poverty. The log likelihood yields a χ^2 of -3692.917 and was significant at ($p < 0.05$) which give a good fit for the model to the data and thus the null hypothesis was also tenable for the model. Values of statistics Cox-Snell R^2 and Nagerlelke R^2 parameters are 0.46 and 0.64 which indicate that the model explains 46% to 64% of the variance in the outcome respectively. The model has a pseudo R^2 of 0.114 which means that 11.3% of the variation in the dependent variable is due to the variations in the independent variables.

Table (4.23) Correct Classification Table [a] of the Model

Observed		Predicted		
		Poverty household status		Percentage Correct
		Non Poor	poor	
Poverty household status	Non poor	4015	45	98.9
	poor	1841	63	3.3
Overall Percentage				68.4

Source: Prepared by the researcher by using SPSS package, 2018

The cut value is .500 [a]

The regression classification table (4.23) revealed that a binary logistic model managed to predict 68.4% of the responses correctly.

Apart from percent correct predictions, the model Chi-Square statistic has been run to evaluate the performance of the model. Accordingly, the Chi-Square value was found to be 4.436 and the overall model was found significant at 0.05 levels.

The classification table (4.23) shows that the model makes a correct prediction of the cases compared to 68.1% in the null model, a marked improvement. Of the 4060 households with non-poor, the model correctly identified 45 households predicted as non-poor by the model are in fact poor. Similarly, of the 1904 households that did have a poor, the model correctly identifies 63 households predicted as poor by the model are in fact non-poor. 3.3% is also known as the **sensitivity** of prediction. 98.9% is also known as the **specificity** of prediction.

The logistic model was fitted to the data to test the relationship between the likelihood of a household being poor or non-poor. The logistic regression analysis was carried out by entering method, and the result showed that the optimal model:

Log $(\frac{P}{1-P})$ =Y= 1.950 - 1.521 CRW (1) -1.608 CRW (2) + .018 No qu - .283 Prim - .274 Intr - .728 Sco - 1.800 Univ -1.415 High + .241 Khl - .169 VT (1) - .276 VT (2) - .142 Nver - 1.195 Mar - .770 Wido - .413 Div - 20.69 Dis (1) - .394 sDis (2)……………………………………….. (4.2)

The model indicates that out of the many variables identified as possible determinants of poverty status only seven were statistically significant. They include; secondary (4), University (5), higher level (6), married (2), widowed (3), divorce (4) disability (2).

In general, the logit model fitted the data quite well. The chi-square test strongly rejects the hypothesis of no explanatory power and the model correctly predicted 68.4% of the observations. Furthermore, the highest level of school ever completed, marital status and Suffering from any type of disability are statistically significant and the sign on the parameter estimate support expectations, while the can read and write with understanding, ever attended vocational training are not significant.

4.7.1.3 Economic Factors

The economic characteristics are discussed and the findings from various respondents presented in terms of cross-tabulation count, by characteristics of the household, like Worked at least one hour for profit in cash or kind last 10 days, dwelling type, main tenure for the dwelling and household main source of livelihood. Further still, the relationship between economic factors and poverty statuses among households are presented by using logistic regression analysis.

Table (4.24) Association between Poverty and Work

			Worked at least one hour for profit in cash or kind last 10 days			Total
			No	Yes	Not stated	
Poverty household status	Non poor	N	701	3343	16	4060
		% within poverty	17.3%	82.3%	0.4%	100.0%
	Poor	N	255	1646	4	1905
		% within poverty	13.4%	86.4%	0.2%	100.0%
Total		N	956	4989	20	5965
		%	16.0%	83.6%	0.3%	100.0%

Source: Prepared by the researcher by using SPSS package, 2018

The results displaying in the table (4.24) above indicate that 13.4% of household heads not working are poor, and 86.4% are work but poor. Majority of the household heads are work and poor; Because of their jobs low-income or due to many dependents in households. Similarly about 17,3% are not working but not poor, and 82.3% are working and not poor.

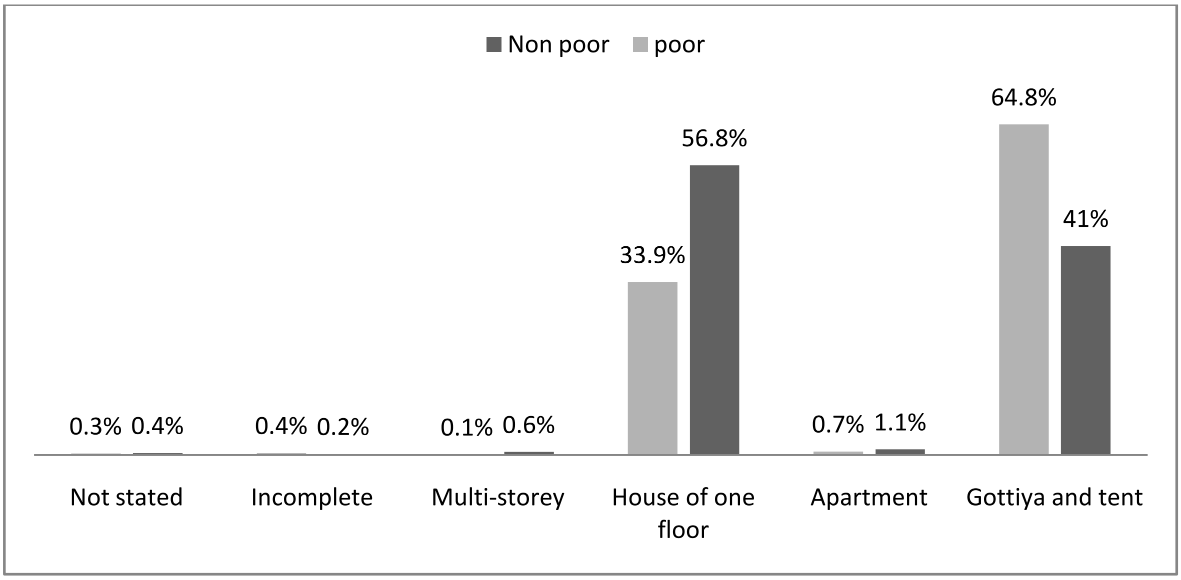
Table (4.25) Relationship between Poverty and Dwelling Type

			Poverty household status		Total
			Non poor	poor	
Dwelling Type (DT)	Gottiya or Tent	Count	1664	1233	2897
		% within poverty	41.0%	64.8%	48.6%
	Apartment	Count	43	13	56
		% within poverty	1.1%	0.7%	0.9%
	House of one floor	Count	2307	645	2952
		% within poverty	56.8%	33.9%	49.5%
	Multi-storey floor	Count	23	1	24
		% within poverty	0.6%	0.1%	0.4%
	Incomplete	Count	8	7	15
		% within poverty	0.2%	0.4%	0.3%
	Not stated	Count	15	5	20
		% within poverty	0.4%	0.3%	0.3%
Total		Count	4060	1904	5964
		% within poverty	100.0%	100.0 %	100.0 %

Source: Prepared by the researcher by using SPSS package, 2018

The distribution of households by Dwelling Type in the table (4.25) above shows that 48.6% of the households lived in Gottiya or Tent. About 64.8% of poor and 41% non poor of households were living in Gottiya or Tent also can see figure (4.5) below. Another 49.5% lived in House of one floor from total respondents, almost 33.9% are poor and 56.8% non poor while 0.7%, 0.1% of households were living in an apartment and multi storey floor respectively are poor. Majority of the houses (50%) had mud/ /brick/concrete to House of one floor while others were made of tent, sticks and straw more obviously in the figure (4.7).

Figure (4.7) association between poverty and dwelling type



Source: Prepared by the researcher by using Excel, 2018

Table (4.26) Relationship between Poverty and Main Tenure Status for the Dwelling

			Main tenure status for the dwelling (MTD)				Total
			Owned	Rented	Housing provided as part of work	Free	
Poverty household status	Non poor	N	3315	263	53	382	4013
		% within poverty	82.6%	6.6%	1.3%	9.5%	100.0%
	poor	N	1514	109	18	239	1880
		% within poverty	80.5%	5.8%	1.0%	12.7%	100.0%
Total		N	4829	372	71	621	5893
		%	81.9%	6.3%	1.2%	10.5%	100.0%

Source: Prepared by researcher by using SPSS package,2018

The distribution of households by Main tenure status for the dwelling (MTD) of residential building as shown in table (4.26) above reveals that 80.5.9% of the household live in owned dwelling are poor's, and only 5.8% rented for poor, while 1.0% of the Housing provided as part of work are poor, while 12.7% were living free. However, 82.6% of the owned and 6.6% household lived in rented are non poor.

Table (4.27) Association between Poverty and Main Source of Livelihood

			Poverty household status		Total
			non poor	poor	
Main sources of livelihood (MSL)	Crop farming	Count	1028	1097	2125
		% within poverty	26.1%	59.6%	36.7%
	Animal husbandry	Count	155	51	206
		% within poverty	3.9%	2.8%	3.6%
	Wages and salaries	Count	1406	316	1722
		% within poverty	35.7%	17.2%	29.8%
	Owned income	Count	873	257	1130
		% within poverty	22.1%	14.0%	19.5%
	Remittances and Aid	Count	132	26	158
		% within poverty	3.3%	1.4%	2.7%
	Transfers from members	Count	349	94	443
		% within poverty	8.9%	5.1%	7.7%
Total		Count	3943	1841	5784
		% within poverty	100.0%	100.0%	100.0%

Source: Prepared by the researcher by using SPSS package, 2018

Table (4.27) indicates that 36.7% and 29.8% of the household's main sources of livelihood earn form crop farming and wages /salaries respectively. The findings also show that about 59.6% of the households earn their means of living exclusively form crop farming's is poor's and only 17.2% receive from wages /salaries, while only 1.4% of the households received main sources of livelihood from remittances and aid were poor. Similarly, about 35.7% of the households received the main sources of livelihood from wages and salaries were non poor.

3- Estimation binary logistic regression for economic variables

The logistic regression technique has been applied to evaluate the economic characteristics of the household's head and household characteristics as the determinants of household poverty in Sudan. The economic factors include worked at least one hour for profit, dwelling type, main tenure status for the dwelling and household main source of livelihood.

Table (4.28) Pearson Chi-square Statistics Test and Likelihood Ratio for the Association between Economic Characteristics and Poverty

Variable	**Pearson Chi-Square**		**Likelihood Ratio**		df
	Value χ^2	P-value	Value χ^2	P-value	
Work	16.052	.000	16.554	.000	2
DT	300.361	.000	306.229	.000	5
MTSD	15.625	.001	15.311	.002	3
HMSL	615.830	.000	609.852	.000	5

Source: Prepared by the researcher by using SPSS package, 2018

The chi-square values of the remaining variables make the null hypothesis to be rejected and the alternative hypothesis to be adopted. The inference is that from the table (4.28), we observe that there is a very strong association between worked at least one hour for profit (work), dwelling type (DT), main tenure for the dwelling (MTSD), the main source of livelihood(HMSL) with Poverty household status.

Table (4.29) Coefficients and Wald Tests for Logistic Regression on the Poverty and Economic Data

	B	S.E	Wald	df	Sig	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1[a] Work			5.063	2	.080			
Work(1)	.370	.604	.375	1	.540	1.447	.443	4.726
Work(2)	.550	.600	.842	1	.359	1.734	.535	5.615
DT			163.021	5	.000			
Got (1)	.929	.539	2.968	1	.085	2.531	.880	7.280
Apa (2)	.151	.631	.057	1	.811	1.163	.338	4.003
Hof(3)	.145	.540	.072	1	.788	1.156	.401	3.330
Mul (4)	-1.323	1.169	1.282	1	.258	.266	.027	2.631
Inc (5)	1.271	.772	2.709	1	.100	3.563	.785	16.179
MTSD			5.202	4	.267			
Own(1)	-.115	.273	.178	1	.673	.891	.521	1.523
Ren(2)	-.039	.299	.017	1	.897	.962	.536	1.728
Hou(3)	-.417	.403	1.074	1	.300	.659	.299	1.450
Free(4)	.072	.286	.063	1	.802	1.074	.613	1.882
MSL			478.166	5	.000			
Cro(1)	1.293	.126	105.273	1	.000	3.644	2.846	4.665
Ana(2)	.195	.202	.928	1	.335	1.215	.818	1.805
WS(3)	-.145	.134	1.179	1	.278	.865	.665	1.124
OI(4)	.085	.138	.379	1	.538	1.089	.830	1.428
TM(5)	-.233	.247	.884	1	.347	.792	.488	1.287
Constant	-2.276	.856	7.077	1	.008	.103		

Source: Prepared by researcher by using SPSS package, 2018

Variable(s) entered on step 1: Work, TD, MTSD, and MSL. a

Table (4.29) shows the results of the logistic model used for possible poverty contributing factors in Sudan. The odds ratios of all variables are more than one, which means that these variables are positively correlated with the probability of being poor. Except six variables (multi-storey floor (4), owned (1), rend (2), housing (3), wages and salaries (3) and transfers from members (5)) all have odds ratios lower than one, which means that these variables are negatively associated with the probability of being poor.

The overall result of the table above shows that out of the four main identified variables only two variables dwelling type (DT) and main sources of livelihood were significant in explaining whether household's status is poor or not poor. The confidence interval for the odd ratios of owned (1), rented (2), housing (3) free (4), animal husbandry (2), Wages and Salaries (3), Owned income (4) and Transfers from members (5) includes the number one, which means that these variables have no statistically significant effect on the probability of poverty. The result shows that the type of dwelling (TD) variables of the households who live in Gottiy and incomplete house were more likely to be poor than lives in Apartment and House of one floor and multistory, this is according to the living situation of the family. Whenever the livelihoods are good, the family needs a suitable house for its situation, so that the type of dwelling is weak relationship with poverty. As for Main tenure status for the dwelling, there is no association with poverty because most families live in their own homes in Sudan. As a result of this, an 82% of Sudan's lives in their homes. The findings show that main sources of livelihood reduce the probability of being poor. From the category of the main sources of livelihood, we find that the crop farming has a statistically significant correlation with poverty, while the rest of the categories have no relation with household poverty status because most of Sudanese are the source of livelihood from agriculture. So that the poverty incidence among crop farmers, because most farmers have a weak capacity of farming. The results indicate that the crop farming positively effects poverty, so it means that when the number of farming households increases, poverty increases in Sudan.

Table (4.30) Omnibus Tests of Model Coefficients

	chi-square	df	sig
Step 1 Step	798.940	16	.000
Block	798.940	16	.000
Model	798.940	16	.000

Source: Prepared by theresearcher by using SPSS package, 2018

In the table (4.30) we have added all four explanatory variables in one block and therefore have only one step. This means that the chi-square values are the same for step, block, and model.

Here the chi-square is highly significant (chi-square=798.940, df = 16, p-value <.000) so our model is significantly better, which indicates the accuracy of the model improves when we add our explanatory variables.

Table (4.31): Summary measures of goodness-of-fit statistics of the model with selected covariates

Summary Statistic test	Value	df	P-value
Hosmer - Lemeshow	6.873	8	.550
-2 Log likelihood	6437.714[a]		
Cox and Snell R^2	.129		
Nagelkerke R^2	.181		

Source: Prepared by the researcher from the Survey Data, 2015

Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.[a]

In order to establish whether the model fits the data Hosmer and Lemeshow (H-L), goodness-of-fit test was undertaken in the table (4.31). The Hosmer and Lemeshow chi-square was 6.873 with 8 degrees of freedom and p-value of test 0.550 indicate that the model fits the data well (P>0.05). Thus we accept the null hypothesis that household characteristics and perceptions have an influence on poverty. Moreover, the shows that the values of Cox and Snell and Nagelkerke R^2 value .129 and .181 respectively indicate that the model is useful in predicting determinants of poverty.

Table (4.32) Correct Classification Table [a] of the Model

Observed		Predicted		
		Poverty household status		Percentage Correct
		Non Poor	poor	
Poverty household status	Non poor	3426	517	86.9
	poor	1080	761	41.3
Overall Percentage				72.4

Source: Prepared by the researcher by using SPSS package, 2018

The cut value is .500[a]

The regression classification table (4.32) revealed that binary logistic model managed to predict 72.4% of the responses correctly.

Apart from percent correct predictions, the model Chi-Square statistic have been run to evaluate the performance of the model. Accordingly, the Chi-Square value was found to be 6.873 and the overall model was found significant at 0.05 levels.

Moreover, the results indicate that, the model makes a correct prediction of the cases compared to 68.2% in the null model, a marked improvement. Of the 3943 households with non-poor, the model correctly identified 517 households predicted as non-poor by the model are in fact poor. Similarly, of the 1841 households that did have a poor, the model correctly identifies 761 households predicted as poor by the model are in fact non-poor. 41.3% is also known as the **sensitivity** of prediction. 86.9% is also known as the **specificity** of prediction.

The logistic model was fitted to the data to test the relationship between the likelihood of a household being poor or non-poor. The logistic regression analysis was carried out by entering method, and the result showed that the optimal model:

Log ($\frac{P}{1-P}$) =Y= - 2.276 + .37o Work (1) + .550 Work (2) + .929 Got(1) + .151 Apa(2) + .145 Hof(3) -1.323 Mul(4) + 1.271 Inc(5) - .115 Own(1) - .039 Ren(2) - .417 Hou(3) + .072 Free(4) + 1.293 Cro(1) + .195 Ana(2) - .145 WS(3)+ .085 OI (4) - .233 TM(5) ……………………………….…….. (4.4)

The model indicates that out of the many variables identified as possible determinants of poverty status only crop farming was statistically significant. In general, the logit model fitted the data quite well. The chi-square test strongly rejects the hypothesis of no explanatory power and the model correctly predicted 72.4% of the observations.

Furthermore, the dwelling type (DT) and main sources of livelihood are statistically significant and the sign on the parameter estimate support expectations, while the main tenure status for the dwelling and worked at least one hour for profit are not significant.

The regression classification table revealed that binary logistic model managed to predict 72.4% of the responses correctly. Apart from percent correct predictions, the model Chi-Square statistic have been run to evaluate the performance of the model. Accordingly, the Chi-Square value was found to be 6.873 and the overall model was found significant at 0.05 levels.

4.7.2 Estimation and Discussion of demography, Social and Economic Factors in one set

In this section the demography and socio-economic characteristics are discussed and the findings from various respondents presented in terms of binary logistic regression analysis, by characteristics of the household, like place of residence, household size, sex of household head, age of household head, dependency ratio of household, can read and write with understanding, highest level of school ever completed, ever attended vocational training, marital status, Suffering from any type of disability that prevents from doing usual work, worked at least one hour for profit, dwelling type, main tenure for the dwelling and main source of livelihood Were used to explain households' poverty.

Table (4.33) Omnibus Tests of Model Coefficients

	chi-square	df	sig
Step 1 Step	2108.352	38	.000
Block	2108.352	38	.000
Model	2108.352	38	.000

Source: Prepared bythe researcher by using SPSS package, 2018

In the above table (4.33) we have added all fourteen explanatory variables in one block and therefore have only one step. This means that the chi-square values are the same for step, block, and model. Here the chi-square is highly significant (chi-square=2108.352, df = 38, p-value <.000) so our model is significantly better, which indicates the accuracy of the model improves when we add our explanatory variables.

Table (4.34) Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	5050.073[a]	.309	.432

Source: Prepared by the researcher by using SPSS package, 2018

a Estimation terminated at iteration number 20

Because maximum iterations has been reached. Final solution cannot be found.

From the table (4.34) the values of statistics Cox-Snell R^2 and Nagerlelke R^2 parameters are 0.309 and 0.432 which indicate that the model explains 31% to 43% of the variance in the outcome respectively. This good value is explained primarily by the fact that the main variable affecting the poor are household income, these variables.

Table (4.35) Hosmer and Lemeshow Test

Step	Chi-Square	df	Sig
1	37.237	8	.000

Source: Prepared by the researcher by using SPSS package, 2018

In order to establish whether the model fits the data Hosmer and Lemeshow (H-L) goodness-of-fit test was undertaken in the table (4.35). The Hosmer-Lemeshow (H-L), test that yields a $\chi^2$of 37.237 and was significant Suggesting that there was lack of fit the model of the data. Thus we rejected the null hypothesis that household characteristics and perceptions have an influence on poverty.

Table (4.36) Correct Classification Table [a] of the Model

Observed		Predicted		
		Poverty household status		Percentage Correct
		Non Poor	poor	
Poverty household status	Non poor	3438	436	88.7
	poor	768	1062	58.0
Overall Percentage				78.9

Source: Prepared by researcher by using SPSS package, 2018

The cut value is .500[a]

The regression classification table (4.36) revealed that a binary logistic model managed to predict 78.9% of the responses correctly. Apart from percent correct predictions, the model Chi-Square statistic has been run to evaluate the performance of the model. Accordingly, the Chi-Square value was found to be 37.237and the overall model was found significant at 0.05 levels.

Moreover, the results indicate that the model makes a correct prediction of the cases compared to 67.9% in the null model, a marked improvement. Of the 3874 households with non-poor, the model correctly identified 436 households predicted as non-poor by the model are in fact poor. Similarly, of the 1830 households that did have a poor, the model correctly identifies 1062 households predicted as poor by the model are in fact non-poor. 58% is also known as the **sensitivity** of prediction. 88.7% is also known as the **specificity** of prediction.

Table (4.37) Coefficients and Wald Test of Logistic Regression for Factors Influencing Household Poverty Data

	B	S.E	Wald	Df	Sig	Exp(B)	95% C.I. for XP(B)	
							Lower	Upper
Step 1[a] PR(1)	.668	.093	51.943	1	.000	1.950	1.626	2.339
HS	.446	.018	648.073	1	.000	1.562	1.509	1.616
SHH(1)	-.306	.109	7.917	1	.005	.736	.595	.911
AHH	-.075	.028	7.126	1	.008	.928	.878	.980
DR	.125	.017	55.707	1	.000	1.134	1.097	1.172
CRW			3.252	2	.197			
CRW(1)	-1.059	1.026	1.067	1	.302	.347	.046	2.588
CRW(2)	-1.267	1.029	1.516	1	.218	.282	.037	2.117
HLS			58.799	7	.000			
No qu(1)	.292	.244	1.425	1	.233	1.339	.829	2.162
Prim(2)	.011	.247	.002	1	.964	1.011	.623	1.642
Inter(3)	-.025	.299	.007	1	.932	.975	.543	1.750
Sco(4)	-.286	.256	1.251	1	.263	.751	.455	1.240
Univ(5)	-1.212	.336	12.984	1	.000	.298	.154	.575
High(6)	-.921	.750	1.508	1	.219	.398	.092	1.731
Khal(7)	.511	.247	4.302	1	.038	1.668	1.029	2.704

VT			4.222	2	.121			
VT(1)	.092	.148	.386	1	.534	1.096	.820	1.465
VT(2)	-.397	.211	3.521	1	.061	.673	.444	1.018
MS			24.295	4	.000			
Never(1)	.098	.945	.011	1	.917	1.103	.173	7.028
Mar(2)	-.996	.316	9.948	1	.002	.369	.199	.686
Wid(3)	-.821	.226	13.134	1	.000	.440	.282	.686
Div(4)	-.376	.259	2.119	1	.145	.686	.413	1.139
Dis			.059	2	.971			
Dis(1)	-19.975	40192.970	.000	1	1.000	.000	.000	.
Dis(2)	-.042	.174	.059	1	.809	.959	.682	1.348
Work			1.178	2	.555			
Work(1)	.653	.708	.850	1	.356	1.921	.480	7.690
Work(2)	.702	.702	.999	1	.317	2.017	.510	7.981
DT			113.683	5	.000			
Got (1)	.835	.663	1.584	1	.208	2.304	.628	8.454
Apa (2)	-.023	.761	.001	1	.976	.977	.220	4.342
Hof(3)	-.030	.664	.002	1	.964	.971	.264	3.570
Mul (4)	-.590	1.265	.218	1	.641	.554	.046	6.615
Inc (5)	.976	.915	1.137	1	.286	2.654	.441	15.960
MTSD			9.834	4	.043			
Own(1)	-.072	.311	.053	1	.818	.931	.506	1.712
Ren(2)	.032	.343	.009	1	.925	1.033	.527	2.022
Hou(3)	-.601	.460	1.708	1	.191	.548	.223	1.350
Free(4)	.230	.326	.496	1	.481	1.258	.664	2.384
MSL			355.927	5	.000			
Cro(1)	1.235	.144	73.194	1	.000	3.439	2.591	4.564
Ana(2)	.267	.230	1.349	1	.245	1.306	.832	2.048
WS(3)	-.238	.152	2.428	1	.119	.789	.585	1.063
OI(4)	.012	.157	.006	1	.938	1.012	.744	1.377
TM(5)	-.302	.283	1.144	1	.285	.739	.425	1.286
Constant	-3.036	1.350	5.059	1	.024	.048		

Source: Prepared by the researcher by using SPSS package, 2018

A binary logistic regression model was run to assess the predictive ability of the selected demography and socio-economic factors on household poverty. According to the summary statistics from the table (4.37) shows that out of the fourteen identified variables only four variable was not significant in explaining whether a household's status is poor or not.

The odds ratios Exp (B) of 21 categories of variables have greater than one, which means that these variables are positively correlated with the probability of being poor. While, as can be seen, 17 category have odd ratios lower than one, which means that these categories are negatively associated with the probability of being poor. The confidence interval for the odds ratios of 27 categories of variables includes the number one, which means that these variables statistically insignificant impact on the probability of poverty.

Type of place of residence PR(1) had significant impact on poverty It had been found that the urban households were less likely to be poor than the rural households (B=0.668, Wald=51.943, p=0.000).

Also, we see that household size (HS) is a statistically significant variable (B= 0.446, Wald = 648.073, P =0.000) and its positive coefficient indicates that with increasing household size increases probability of that household be poor.

Thus, even though the coefficient (B= -0.306, Wald=7.917, P-value=.005) for the Sex of household head SHH (1) variable is negative and statistical significance. The gender of the head variable is an important factor in explaining the poverty status of the family but the negative coefficient indicates that households headed by a female have the lower probability of being poor than male-headed households.

The coefficient for the variable age of household head (AHH) is negative and statistically significant variable (B= -0.075, Wald= 7.126. P=.008). This means that there is an established negative relationship between age of the household head and the per capita expenditure of the household. The age of household heads grows older; the per-capita expenditure/income of the household reduces, thus, increasing the level of poverty in the household. We found that there is a strong and statistically significant inverse relationship between poverty and age of household head.

The coefficient of the dependent ratio (DR) is positive and statistically significant (B=.125, Wald= 55.707, P= .000). The odds ratio of the variable dependency ratio shows a contribution of 17.2% in increasing the likelihood of being poor whereas household size (HS) contributes 61.6%. Therefore a majority of households fell into poverty because of having large families with many dependants being children or elderly at unproductive age.

The study shows that the variable can read and write with understanding does not a significantly affect poverty level probably (Wald=3.252, P-value=.197) and negative relation with the probability of being poor, because of lower education have more chances to be poor. The can read and write with understanding is not significant impact household poverty.

Levels of education had been detected to significantly explain the poverty of different households (Wald=58.799, p-value=0.000). In this context, the results depict that there was a negative relationship between the probability of being a poor and level of education such as; intermediate, secondary, High and university. While the no qualification, primary and khalowa were a positive association with the probabilities of being a poor of household. It means that higher levels of education reduce the probability of being poor gradually. If we look at the no qualification, primary, intermediate, post-graduate and secondary are not statistically significant. But the university and khalowa remain an important determinant of household welfare and say are statistically significant variable. The university level negative coefficients (B= -1.212, W=12.984, p=.000) indicate that increased education has a significant impact in reducing the probability of being poor. While the khalowa positive coefficient (B=.511, Wald=4.302 p=.o38) means that increased khalowa a significant affected in increasing the probability of being poor. The odds ratio of the variable khalowa shows a contribution of 70% in increasing the likelihood of being poor .This implies that education is the important factors in reducing the impact of poverty at the household level. These findings confirmed the conclusions of other studies, such as Bigsten et al. (2003); Achia, (2010), Sarwar et al (2012) and Xhafaj, & Nurja, (2014).

The results also indicate that the vocational training (Wald=4.222, P-value=.121) is not significant in explaining the probability of being poor, this was due to lack of adequate training in Sudan.

Furthermore, the marital status is a statistically significant variable (Wald =24.295, P-value=.000) and negatively correlated with responsiveness excepted never married positively correlated with poverty. This is due to the presence of family solidarity system in Sudan. Moreover, the study shows that the married household heads have a higher chance of being poor as compared to household heads that are not married. More specifically, the results indicate that the married and widowed head of households were significantly impacted to be poor. This may be as a result of having more dependants depending on the household head. We also found that the never married and divorce insignificant effect on the probability of being poor.

Moreover, we found that the Suffering type of disability is not significant (Wald = .059, P-value=.971) in explaining the probability of being a poor and negative relation with a probability of being poor.

The study shows that the Worked at least one hour for profit (work) is insignificant impact with probability of being poor household (Wald=1.178, p=.555).

The dwelling types (DT) (Wald=113.683, p-value=.000) was significant in explaining whether household's status is poor or not. The result shows that the type of dwelling (TD) category the households who live in Gottiy and incomplete house were positive coefficient and insignificant. But the lives in Apartment, House of one floor and multistory floor are negative coefficient and not significant with poverty.

As for Main tenure status for the dwelling (MTSD) (Wald=9.834, p-value=.043), there is no association with poverty because most families live in their own homes in Sudan. As a result of this, 82% of Sudan's lives in their homes. The findings show that the main sources of livelihood reduce the probability of being poor.

From the category of the main sources of livelihood (MSL) (Wald=355.927, p-value=.000), we find that the crop farming has a statistically significant correlation with poverty (B=1.235, Wald=73.194, p=.000), while the rest of the categories have no relation with household poverty status, because most of Sudanese are the source of livelihood from agriculture. The results indicate that the crop farming positively effects poverty, so it means that when the number of farming households increases, poverty increases this for traditional agriculture in Sudan. The odds ratio of the variable crop farming shows a contribution of 56% in increasing the likelihood of being poor so that the poverty incidence among crop farmers in Sudan.

The logistic model was fitted to the data to test the relationship between the likelihood of a household being poor or non-poor. The logistic regression analysis was carried out by entering method, and the result showed that the optimal model:

Log($\frac{P}{1-P}$)=Y= - 3.036 + .668PR(1) + .446HS - .306SHH(1) - .075AHH + .125DR - 1.059CRW(1) - 1.267CRW(2) + .292NO Qu(1) + .011Prim(2) - .025Inter(3) - .286Sco(4) - 1.212Univ(5) - .921High(6) + .511Khal(7) + .092VT(1) - .397VT(2) + .098Never(1) - .996Mar(2) - .821Wid(d) - .376Div(4) - 19.975Dis(1) -.042Dis(2) + .653Work(1) + .702Work(2) + 835Got(1) - .023Apa(2) - .030Hof(3) - .590Mul(4) + .976Inc(5) - .072Own(1) + .032Ren(2) -.601Hou(3) + .230Free(4) + 1.235Gro(1) + .267Ana(2) - .238WS(3) + .012OI(4) - .302TM ……………… (4.6)

The model indicates that out of the many variables identified as possible determinants of poverty status only ten were statistically significant. They include; place of residence (PR), household size (HS), sex of household head (SHH), the age of household head (AHH), dependency ratio (DR), University (5), khalowa (7), married (2), widowed (3), and agriculture (1).

P = $\frac{e^{a+B_1X_1+B_2X_2+\ldots+B_kX_k}}{1+e^{a+B_1X_1+B_2X_2+\ldots+B_kX_k}}$, o$< p < 1$ …..………… (4.7)

The estimates of the logistic regression are shown in the above Tables. In general, the logit model fitted the data quite well. The chi-square test strongly rejects the hypothesis of no explanatory power and the model correctly predicted 78.9% of the observations.

Finally, according to the hypotheses of the study, we found that; there is a significant relationship between the demography factors and the poverty. While the social factors, there are only University and khalow from category (HLS) and married and widowed from category

(MS) a significant association with poverty. Moreover, the economic factors there are only crop farming from category (MSL) a significant relationship with household poverty status.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1Preface

5.2 Conclusions

5.3 Recommendations

CONCLUSIONS AND RECOMMENDATIONS

5.1 Preface

This study has tried to address the estimation and analysis of factors affecting poverty across selected households in Sudan. This chapter presents a summary of the findings, in-depth discussions and conclusion that poverty and households can get valuable information from, and further recommendations of the study are found here too.

5.2 Conclusions

Therefore we can conclude from the results reported above that:

1- The average of the household size of poor was found to be 7.37 and non poor households were 5.12. While the minimum and the maximum number of members in the family for poor households was found to be 20.2 members and for non poor was 17.1 sizes respectively.

2- The average dependency ratio of households for poor and non poor was 163.05 and 99.5 percent respectively.

3- The results show that 31.1% of the more poorest households are aged between 35 and 44 year, as compared to 25.47% of the poor households head who are over 54 year and 16.28% of the poor who are less than 35 years, while 58.25% of the poor’s between 35 and 54 years, and more poor household head in age 40.

4- The value of statistic Nagelkreke R^2 parameters is 0.432 which indicate that the model explains 43.2% of the variance in the outcome.

5- The chi-square test strongly rejects the hypothesis of no explanatory power and the model correctly predicted 78.9% of the observations.

6- The result reveals that place of residence, household size, sex of household head, the age of household head, dependency ratio, University, khalowa, married, widowed, and crop farming significantly explains the poverty status of a household.

7- The place of residence, household size, dependency ratio, khalowa and agriculture were positively associated with the probability of being poor.
8- The sex of household head, the age of household head, university, married and widowed was negatively related to the poverty status.

5.3 Recommendations

The analysis undertaken in this study leads to the following guidelines implications for the researcher and government.

- ✓ The study recommends using the optimal model of the logistic regression to predict household poverty in future through the variables that affect it.
- ✓ The study recommends government and households should be focusing on the significances variables found in this study and highlight the mechanisms that can be beneficial in the reduction of the poverty levels in Sudan.
- ✓ This study recommends a careful review on the reforms to be taken in relation to household size and dependency ratio, suggests that an intensification of family planning programmed at the grass root level are amongst rural areas.
- ✓ Poverty alleviation efforts should be made to improve the socio-economic and demographic characteristics of the households in general and demographic factors in particular since the number of the poor is increased in both urban and rural areas. The government should be focusing on improving the livelihood situation.
- ✓ The government should look at the labor conditions of females and to reduce poverty, great attention must be paid to the manufacturing sector and agriculture.
- ✓ This study recommends a careful review of the reforms to be taken in relation to education and poverty, suggests that expansion of education and vocational training programmed at the grass root level are amongst rural areas.
- ✓ Future research is needed to yield more results on poverty predictors in Sudan and to determine where poor communities have actually been lifted out of the poverty trap. The variables

used in this study include some of the most important predictors of poverty, but other variables also have an impact on poverty such as; health care access.

- ✓ Future research could look at the significant variables found in this study and seek to analyze in greater depth the channel through which these variables interact with poverty and also there is need of more recent data to capture the recent trends in poverty.
- ✓ Results from this study revealed that in future, the poverty may be predicted by considering these identified influential variables. We recommend using logistic regression to measure the impact of different poverty factors by utilizing per capita expenditures particularly in developing countries.
- ✓ Logistic regression was used in this study to identify the key determinants of poverty in Sudan. Although this was the best model with this kind of dependent variable, I feel there are other models such as; neural network models that can be used to produce a more accurate output for this study. I strongly recommend other researchers to try this.

References

REFERENCES

[1] Abaker, A.M. and Salih, A.A., 2012, April. Poverty and inequality in Services Sector of Sudan. IJPSS Volume 2, Issue 4.

[2] Abd Razak, N. Ibrahim. N; 2014.Determinants of poverty in Malysia. Journal of Social Sciences Research ICSSR 2014 (e-ISBN 978-967-11768-7-0). 9-10 June 2014, Kota Kinabalu, Sabah, MALAYSIA.

[3] ABUSIN, A. (2003), IPRSP- Sudan (ECA-LG). Addis Ababa: The economic commission for Africa.

[4] Achia, T.N., Wangombe, A. and Khadioli, N., 2010. A logistic regression model to identify key determinants of poverty using demographic and health survey data. European Journal of Social Sciences, 13(1), pp.38-45.

[5] Adetayo, A., 2014. Analysis of farm households poverty status in Ogun states, Nigeria. Asian Economic and Financial Review, Vol. 4, 3, 325-340.

[6] Ahmed, A.E., Roghim, S. and Saleh, A; 2011. Poverty Determinants in South Sudan: The Case of Renk County.

[7] Akerele, D. and Adewuyi, S.A., 2011. Analysis of poverty profiles and socioeconomic determinants of welfare among urban households of Ekiti State, Nigeria. Current Research Journal of Social Sciences, 3(1), pp.1-7.

[8]Akerele, D., Momoh, S., Adewuyi, S.A., Phillip, B.B. and Ashaolu, O.F., 2012. Socioeconomic determinants of poverty among urban households in South-West Nigeria. International journal of social economics, 39(3), pp.168-181.

[9] Ali, A.A.G., 2002. Building human capital for economic development in the Arab countries. Egyptian Center for Economic Studies.

[10] Alkire, S., Foster, J., Seth, S., Santos, M., Roche, J. and Ballon, P., 2015. Multidimensional poverty measurement and analysis: Chapter 5–the Alkire-Foster counting methodology.

[11]Asad Khan Sept 2010, September. Model Building using Logistic Regression. Australia University of Queensland.

[12] Awad, M.H., 1997. Poverty in Sudan: Anatomy and prognosis. In UNDP Workshop on Poverty in Sudan, Sharja Hall (14-15 june), Khartoum.

[13] Awan, M.S., Malik, N., Sarwar, H. and Waqas, M., 2011. Impact of education on poverty reduction.

[14] Bahta, Y.T. and Haile, B.O., 2013. Determinants of Poverty of Zoba Maekel of Eritrea: A Household Level Analysis. International Journal of Food and Agricultural Economics, 1(2), pp.73-84.

[15] Balarabe, I.I., 2014. Empirical Investigation of the Determinants of Poverty in Kano Metropolis, Nigeria. Journal of Economics and Sustainable Development Vol.5, No.14, 2014.

[16] Ballon, P. and Duclos, J.Y., 2015. Multidimensional Poverty in Sudan and South Sudan.

[17] Bigsten, A., Kebede, B., Shimeles, A. and Taddesse, M., 2003. Growth and poverty reduction in Ethiopia: Evidence from household panel surveys. World development, 31(1), pp.87-106.

[18] Blundell, R.W. and Powell, J.L., 2004. Endogeneity in semiparametric binary response models. The Review of Economic Studies, 71(3), pp.655-679.

[19] Bogale, A., Hagedorn, K. and Korf, B., 2005. Determinants of poverty in rural Ethiopia. Quarterly Journal of International Agriculture, 44(2), pp.101-120.

[20] Bolin, N., 2014. New party parliamentary entry in Western Europe, 1960-2010. European Journal of Government and Economics, 3(1), pp.05-23.

[21] Brück, T., Danzer, A.M., Muravyev, A. and Weifphaar, N., 2008. Determinants of poverty during transition: household survey evidence from Ukraine.

[22] Bukhari, S.Z., Central Line Associated Blood Stream InfecΘon Rate after IntervenΘon and Comparing Outcome with NaΘonal Healtcare Safety Network and InternaΘonal Nosocomial InfecΘon Control ConsorΘum Data.

[23] Burns, R.P. and Burns, R., 2008. Business research methods and statistics using SPSS. Sage.

[24] Busisa, E., 2011. Determinants of poverty among households in Uganda (Doctoral dissertation, Makerere University).

[25] Cameron, A.C. and Trivedi, P.K., 2005. Microeconometrics: methods and applications. Cambridge university press.

[26] Carter, D.B. and Signorino, C.S., 2010. Back to the future: Modeling time dependence in binary data. Political Analysis, 18(3), pp.271-292.

[27] Central bureau of statistics, poverty in northern Sudan, Household Survey data, 1992, Khartoum.

[28] Central bureau of statistics, poverty in northern Sudan, the estimates National Baseline Household Survey data, 2009, Khartoum, 2010.

[29] Central bureau of statistics,2003. Sudan, the estimates Sudan Health Situation analysis, 2003, Khartoum, 2004.

[30] Chaudhry, I.S., Malik, S. and ul Hassan, A., 2009. The impact of socioeconomic and demographic variables on poverty: a village study. The Lahore Journal of Economics, 14(1), p.39.

[31] Chen, S. and Ravallion, M., 2013. More relatively-poor people in a less absolutely-poor world. Review of Income and Wealth, 59(1), pp.1-28.

[32] Chromy, J.R. and Abeyasekera, S., 2005. Chapter 19: Statistical analysis of survey data. Household Sample Surveys in Developing and Transition Countries. United Nations, Department of Economic and Social Affairs.
[33] Cokluk, O., 2010. Logistic Regression: Concept and Application. Educational Sciences: Theory and Practice, 10(3), pp.1397-1407.
[34] Coudouel, A., Hentschel, J.S. and Wodon, Q.T., 2002. Poverty measurement and analysis. A Sourcebook for poverty reduction strategies, 1, pp.27-74.

[35] Cox, David Roxbee, and E. Joyce Snell. Analysis of binary data. Vol. 32. CRC Press, 1989.
[36] Cragg, J.G. and Uhler, R.S., 1970. The demand for automobiles. The Canadian Journal of Economics/Revue Canadienne d'Economique, 3(3), pp.386-406.
[37] Cressie, N. and Chan, N.H., 1989. Spatial modeling of regional variables. Journal of the American Statistical Association, 84(406), pp.393-401.
[38] De Haan, A., 2000. Social Exclusion: Enriching the understanding of deprivation. Studies in Social and Political Thought, 2(2), pp.22-40.

[39] Deaton, A. and Zaidi, S., 2002. Guidelines for constructing consumption aggregates for welfare analysis (Vol. 135). World Bank Publications.
[40] DFID (2005), Sudan country engagement plan. London: Department for International
Development (DFID).

[41] Eberhardt, L.L. and Breiwick, J.M., 2012. Models for population growth curves. ISRN ecology, 2012.
[42] EL-BATTHANI, A., A. KARIB. & I. SAID (1998), Study of urban problems. Faculty of Economic and Political Science, University of Khartoum, Sudan.

[43] ELTIGANI, E.E.T., ed. (1995), War and drought in Sudan: Essays on population displacement. Gainesville: University Press of Florida.

[44] Englama, A. and Bamidele, A., 1997, September. Measurement issues in poverty. In Poverty Alleviation in Nigeria, Selected Papers for the 1997 Annual Conference of the Nigerian Economics Society (pp. 141-156). Ibadan: Nigerian Economics Society.

[45] Farah, N., 2015. Impact of Household and Demographic Characteristics on Poverty in Bangladesh: A Logistic Regression Analysis.

[46] Fofack, H., 2002. The nature and dynamics of poverty determinants in Burkina Faso in the 1990s (Vol. 2847). World Bank Publications.

[47] Foster, J., Greer, J. and Thorbecke, E., 1984. A class of decomposable poverty measures. Econometrica: Journal of the Econometric Society, pp.761-766.

[48] Foster, J., Greer, J. and Thorbecke, E., 2010. The Foster–Greer–Thorbecke (FGT) poverty measures: 25 years later. The Journal of Economic Inequality, 8(4), pp.491-524.

[49] Garza-Rodriguez, J., 2015. The determinants of poverty in the Mexican states of the US-Mexico border. Browser Download This Paper.

[50] Geda, A., de Jong, N., Kimenyi, M.S. and Mwabu, G., 2005. Determinants of poverty in Kenya: a household level analysis.

[51] George, E.I., 2000. The variable selection problem. Journal of the American Statistical Association, 95(452), pp.1304-1308.

[52] Gerolimetto, M., Mauracher, C. and Procidano, I., Poverty and Transition: the Case of Bulgaria.

[53] Giancristofaro, R.A. and Salmaso, L., 2007. Model performance analysis and model validation in logistic regression. Statistica, 63(2), pp.375-396.

[54] Gounder, N., 2012. The determinants of household consumption and poverty in Fiji. Griffith University.

[55] Gounder, Neelesh. "The Determinants of Household Poverty in Fiji." 2011.

[56] Habyarimana, F., Zewotir, T. and Ramroop, S., 2015. Analysis of demographic and health survey to measure poverty of household in Rwanda. Etude de la Population Africaine, 29(1), p.1472.

[57] Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L., 2006. SEM: confirmatory factor analysis. Multivariate data analysis. Pearson Prentice Hall, Upper Saddle River, pp.770-842.

[58] Hallo Zubair. 2006. Human poverty index account. University of Sudan, 2006, 5.

[59] Hamid Faki, Eltahir M. Nur and Abdelaziz Hashim (ICARDA, forthcoming). Poverty assessment and mapping in the Sudan (Final Draft 2010). Study prepared in cooperation with ICARDA and IFAD.

[60] Hari Srinivas; 2015. Causes of Poverty, A GDRC compilation of articles on the causes of poverty and related issues. Sourcebook Series E-093. January 2015.

[61] Harrell, F.E., Lee, K.L. and Mark, D.B., 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in medicine, 15(4), pp.361-387.

[62] Hentschel, J. and Lanjouw, P., 1996. Constructing an indicator of consumption for the analysis of poverty: principles and illustrations with reference to Ecuador (Vol. 124). World Bank Publications.

[63] Herron, M.C., 1999. Postestimation uncertainty in limited dependent variable models. Political Analysis, 8(1), pp.83-98.

[64]Hilbe, J.M., 2011. Logistic regression. In International Encyclopedia of Statistical Science (pp. 755-758). Springer Berlin Heidelberg.

[65] Homser, D.W. and Lemeshow, S., 1989. Applied logistic regression. New York (NY): J Wiley and Sons.

[66] Hosmer Jr, D.W., Lemeshow, S. and Sturdivant, R.X., 2013. Applied logistic regression (Vol. 398). John Wiley & Sons.

[67] Ibrahim, Badr-El-Din A. (2003). Poverty Alleviation via Islamic Banking Finance to Micro-Enterprises (MEs) in Sudan: Some lessons for poor countries. , Sudan Economy Research Group, University of Bremen, Germany

[68] International Monetary Fund (IMF). 2013. IMF Country Report No. 13/318t: Sudan.

[69] Ismael Mohammed Daffallah. 2001. Economic Growth, Resource Allocation and Poverty Reduction in Sudan (1990-2000), Sudan, 2001, 1.

[70] James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. Linear model selection and regularization. In An Introduction to Statistical Learning (pp. 203-264). Springer New York.

[71] Justino, P. and Litchfield, J., 2003. Poverty dynamics in rural Vietnam: winners and losers during reform. Poverty Research Unit at Sussex Working Paper, 10.

[72] Kamgnia, B.D. and Timnou, J.P., 2008. The determinants of poverty in Cameroon. Developing a sustainable economy in Cameroon.

[73] Khudri, M.M. and Chowdhury, F., 2013. Evaluation of socio-economic status of households and identifying key determinants of poverty in Bangladesh. European Journal of Social Sciences, 37(3), pp.377-387.

[74] Kleinbaum, D.G. and Klein, M., 2010. Analysis of matched data using logistic regression. In Logistic regression (pp. 389-428). Springer New York.

[75] Kochel, T.R., Wilson, D.B. and Mastrofski, S.D., 2011. EFFECT OF SUSPECT RACE ON OFFICERS'ARREST DECISIONS. Criminology, 49(2), pp.473-512.

[76] Kong, F. and Fao, G., 2005. Hypothesis on cyanobacteria bloom-forming mechanism in large shallow eutrophic lakes. Acta ecologica sinica/Shengtai Xuebao, 25(3), pp.589-595.

[77] KOSSAIFI, G. F. (1998), Poverty in the Arab world: Toward a critical approach. Paper presented at the Mediterranean Development Forum, September 3-6, Marrakech, Morocco.

[78] Kotze, D.J., Zeeman, L., Niehaus-Coetzee, E. and Roux, J.P., 2014. Back@ Work: Managing Change Following Unprotected Industrial Action in the Mining Industry.

[79] Lee, L.F., 1983. Generalized econometric models with selectivity. Econometrica: Journal of the Econometric Society, pp.507-512.

[80] Lesaffre, E. and Albert, A., 1989. Multiple-group logistic regression diagnostics. Applied Statistics, pp.425-440.

[81] Lipton, M., 1983. Demography and poverty.

[82] Longford, N.T., 1994. Logistic regression with random coefficients. Computational Statistics & Data Analysis, 17(1), pp.1-15.

[83] Lerman, R.I., 2002. Impacts of Marital Status and Parental Presence on the Material Hardship of Families with Children.

[84] Maddala, G.S., 1983. Limited-dependent and qualitative variables in econometrics (No. 3). Cambridge university press.

[85] Magrabi, F.M., 1991. The economics of household consumption. Greenwood Publishing Group.

[86] Makame, I.H. and Mzee, S.S., 2014. Determinants of Poverty on Household Characteristics in Zanzibar: A logistic Regression Model. Developing Country Studies, Vol.4, No.20, 2014.

[87] Ministry of social welfare and women and Children Affairs. 2004. Definition and measurement of poverty and efforts to combat it, Khartoum, 2004.

[88] Mohammed, B. M; 2017. Measurement and Determinants of Urban Poverty in Case of Southern Nations, Nationalities, and Peoples' Region (SNNPR), Ethiopia. International Journal of Scientific and Research Publications, Volume 7, Issue 3, March 2017, 181 .ISSN 2250-3153.

[89] Mubasher, U., 2009. Socioeconomic Determinants of Poverty A Case of Pakistan. Development and International Relations, Aalborg University, Denmark, 10th Semester 2009.

[90] Myers, R.H., Montgomery, D.C., Vining, G.G. and Robinson, T.J., 2012. Generalized linear models: with applications in engineering and the sciences (Vol. 791). John Wiley & Sons.

[91] Nataša, P, B; 2016. POVERTY ANALYSIS USING MACHINE EARNING METHODS, BACHELOR THESIS., COMENIUS UNIVERSITY IN BRATISLAVA FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS, 2016.

[92] Nelder, J.A. and Baker, R.J., 1972. Generalized linear models. John Wiley & Sons, Inc.

[93] Ngunyi, A., 2015. Multidimensional Analysis of the Determinants of Poverty Indicators in the Lake Victoria Basin (Kenya). IOSR Journal of Mathematics (IOSR-JM). V (May - Jun. 2015), PP 73-87.

[94] O'Hare, G. and Rivas, S., 2007. Changing poverty distribution in Bolivia: the role of rural–urban migration and urban services. GeoJournal, 68(4), pp.307-326.

[95] Obwona, M. and Guloba, M., 2009. Poverty reduction strategies during post-conflict recovery in Africa. Journal of African Economies, 18(suppl_1), pp.i77-i98.

[96] Ogwumike, E., 1998. Structural adjustment and poverty in Africa. In Conference on African perspectives on structural adjustment. Council for the Development of Social Science Research in Africa, Abidjan, Cote d'lvoire.

[97] Osinubi, T.S., 2005. Macroeconometric analysis of growth, unemployment and poverty in Nigeria. Pakistan Economic and Social Review, pp.249-269.

[98] Pan, W., 2001. Model selection in estimating equations. Biometrics, 57(2), pp.529-534.

[99] Park, H., 2013. An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. Journal of Korean Academy of Nursing, 43(2), pp.154-164.

[100] Peng, C.Y.J. and So, T.S.H., 2002. Logistic regression analysis and reporting: A primer. Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences, 1(1), pp.31-70.

[101] Pregibon, D., 1981. Logistic regression diagnostics. The Annals of Statistics, pp.705-724.

[102]Puurbalanta, R. and Adebanji, A.O., 2016. Household Poverty-Risk Analysis and Prediction Using Bayesian Ordinal Probit Models. International Journal of Statistics and Applications, 6(6), pp.399-407.

[103] Ravallion, M., 1996. Issues in measuring and modeling poverty.

[104]Ravallion, M., 1998. Transient poverty in postreform rural China. Journal of Comparative Economics, 26(2), pp.338-357.

[105] Read, T.R. and Cressie, N.A., 2012. Goodness-of-fit statistics for discrete multivariate data. Springer Science & Business Media.

[106] Reddy, O.C. and Alemayehu, E., 2015. Ordinal logistic regression analysis to assess the factors that affect health status of students in Ambo University: a case of natural and computational sciences college, Ambo University.

[107] Ron Heck, S; 2012. Seminars; Introducing Generalized Linear Models: Logistic Regression. Application Seminar.

[108] Rossiter, D.G. and Loza, A., 2008. Technical Note: Analyzing land cover change with logistic regression in R. ITC, Enschede, The Netherlands.

[109] Rowntree, B.S., 1901. Poverty: A study of town life. Macmillan.

[110] Sabir. H. M., February 2012.The Impact of Different Macroeconomic variables on Poverty in Pakistan. Journal of contemporary research in business, p3,10.

[111] Sachs, J.D. and McArthur, J.W., 2005. The millennium project: a plan for meeting the millennium development goals. The Lancet, 365(9456), pp.347-353.

[112] Sackey, H.A., 2005. Female labour force participation in Ghana: The effects of education (Vol. 150). African Economic Research Consortium.

[113] Sakuhuni, R.C., Chidoko, C., Dhoro, N.L. and Gwaindepi, C., 2011. Economic Determinants of Poverty in Zimbabwe. International Journal of Economic Research, 2(6), pp.1-12.

[114] Sancho, J.V., Hernández, F., López, F.J., Hogendoorn, E.A. and Dijkman, E., 1996. Rapid determination of glufosinate, glyphosate and aminomethylphosphonic acid in environmental water samples using precolumn fluorogenic labeling and coupled-column liquid chromatography. *Journal of chromatography A*, *737*(1), pp.75-83.

[115] Sekhampu, T.J., 2013. Determinants of poverty in a South African township. Journal of Social Sciences, 34(2), pp.145-153.

[116] Semasinghe, W.M., 2009. Social development of Sri Lanka: achievements and challenges.

[117] Sen, A., 1976. Poverty: an ordinal approach to measurement. Econometrica: Journal of the Econometric Society, pp.219-231.

[118] Sikander, M.U. and Ahmed, M., 2008. Household Determinants of Poverty in Punjab: A logistic Regression Analysis of MICS (2003-04) data set. In 8th global Conference on Business and Economics.

[119] Sindzingre, A., 2000. Exclusion and Poverty in Developing Countries. In Koherdorfer- Lucius G. and Pleskovic B.(eds.) Inclusion Justice and Poverty Reduction Villa Bosig Workshop Series 1999. German Foundation of International Development. Berlin.

[120] Sinnathurai, V. and Březinová, O., 2012. Poverty Incidence and its determinants in the estate sector of Sri Lanka. Journal of Competitiveness.

[121] Smith, A. (1776). The Wealth of Nations. Ch. VIII P.94.

[122] Spaho, A., 2014. Determinants of Poverty in Albania. Journal of Educational and Social Research, 4(2), p.157.

[123] Stephenson, B., Cook, D., Dixon, P., Duckworth, W., Kaiser, M., Koehler, K. and Meeker, W., 2008. Binary response and logistic regression analysis. available at:< a href=" http://www. stat. wisc. edu/mchung/teaching/MIA/reading/GLM. logistic. Rpackage. pdf"> http://www. stat. wisc. edu/mchung/teaching/MIA/reading/GLM. logistic. Rpackage. pdf(last access: 30 August 2014).

[124] Torres-Reyna, O., 2012. Getting started in Logit and ordered logit regression. Princeton University, http://dss. princeton. edu/training/Logit. pdf.

[125] Tranmer, M. and Elliot, M., 2008. Binary logistic regression. Cathie Marsh for census and survey research, paper, 20.

[126] Tuyen, T.Q., 2015. Socio-Economic Determinants of Household Income among Ethnic Minorities in the North-West Mountains, Vietnam. Croatian Economic Survey, 17(1), pp.139-159.

[127] United Nations. (2005). Indicators of Poverty and Hunger. [pdf]. Retrieved from.http://www.un.org/esa/socdev/unyin/documents/ydiDavidGordon_poverty.pdf.

[128] Usman, M. and DIR, M.S., 2009. A Case of Pakistan Development and International Relations, Aalborg University, Denmark, 10th Semester 2009

[129] Williams, M. and Burden, R.L., 2004. Psycology for language teachers. Ernst Klett Sprachen.

[130] Williamson, J., 2000. What should the World Bank think about the Washington Consensus?. The World Bank Research Observer, 15(2), pp.251-264.

[131] World Bank. 2001. World development report2000/2001. New York: Oxford University Press.

[132] World Bank., 1997. Research Observer, master list of articles. Available at http://www.worldbank.org/research/journals/wbromast.htm.

[133] World Health Organization. Violence, Injury Prevention and World Health Organization, 2013. Global status report on road safety 2013: supporting a decade of action. World Health Organization.

[134] Wright, R.E., 1995. Logistic regression.

[135] Wu, H. and Zhang, J.T., 2006. Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches (Vol. 515). John Wiley & Sons.

[136] Waite, L. and Gallagher, M., 2000. The case for marriage: Why married people are healthier, happier, and better-off financially. *Westminster, MD: Broadway Books*.

[137] Xhafaj, E. and Nurja, I., 2014. Determination of The Key Factors That Influence Poverty Through Econometric Models. European Scientific Journal, ESJ, 10(24).

[138] Yusyf, H. M., 2015. Determinants of Rural Poverty in Tanzania. ISSN 2224-607X (Paper) ISSN 2225-0565 (Online) Vol.5, No.6, 2015.

[139] Zyka, E. and Bici, R., 2014. Identifying Household Level Determinants of Poverty in Albania Using Logistic Regression Model.

[140] http://www.medicine.mcgill.ca/epidemiology/Joseph/courses/EPIB-621/logselect.pdf.

[141] http://www.stat.columbia.edu/~martin/W2024/R10.pdf.

[142] http://www.medicine.mcgill.ca/epidemiology/Joseph/courses/EPIB-621/logselect.pdf.

[143] (n.a) Estimation and hypothesis testing for logistic regression; February 19, 2004 ().https://courses.washington.edu/b515/l13.pdf

[144] (n.a) STATISTICA Formula Guide: Logistic Regression2013. http://documentation.statsoft.com/portals/0/formula%20guide/Logistic%20Regression%20Formula%20Guide.pdf.

[145] (n.a) Binary Response Models. hsttp://jdemeritt.weebly.com /uploads/2/2/7/7/22771764/binaryresponse.pdf.

[146] https://ourworldindata.org/extreme-poverty

[147] http://www.foxbusiness.com/markets/2012/09/14/10-poorest-countries-in-world.html Sudan Overview - World Bank Group http://www.worldbank.org/en/country/sudan/overview.

[148]http://documentation.statsoft.com/portals/0/formula%20guide/Logistic%20Regression%20Formula%20Guide.pdf. STATISTICA Formula Guide: Logistic Regression2013.

[149]http://pegasus.cc.ucf.edu/~zhang/STA6938/topic4.pdf.STA6938-Logistic Regression Model-Topic 4-Model Assessment.

[]https://stats.idre.ucla.edu/other/mult-kg/faq/general/faqhow-are-the-likelihood-ratio-wald-and-lagrange-multiplier-score-tests-different-andor-similar/.

[150] Gerolimetto. M, Advanced data analysis LOGISTIC REGRESSION

www.dst.unive.it/~margherita. without date.s