EFFECT OF SOCIOECONOMIC AND DEMOGRAPHIC FACTORS ON THE HOUSEHOLD POVERTY IN SRI LANKA: A LOGISTIC REGRESSION APPROACH

L. U. KALPAGE

(158378A)

Thesis submitted in partial fulfillment of the requirements for the degree Master of Science

Department of Mathematics

University of Moratuwa

Sri Lanka

January 2020

Declaration of the Candidate and the Supervisor

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the Masters thesis under my supervision.

Signature of the supervisor: Date:

Abstract

Poverty exists when people lack to satisfy their basic needs. To address the household poverty, it is needed to determine the basic needs of a household. This may be defined as narrowly as "those necessary for survival" or as broadly as "those reflecting the prevailing standard of living in the community". Although Sri Lanka has downward trend in poverty, still considerable number of households are poor. Therefore, this study was trying to identify the determinants of the poverty of households in Sri Lanka.

The major objective of this study is to identify the socioeconomic and demographic factors that mainly associated with household poverty in Sri Lanka. To accomplish the objective, Logistic Regression Model is used.

Data gathered for the current study from Household Income and Expenditure survey (HIES) – 2016 conducted by the Department of Census and Statistics in Sri Lanka. According to the descriptive analysis of sample data of the HIES, 2.1% poor households are present. Out of these poor households, most of the poor households are in Batticaloa (10.5%).

According to the results of Binary Logistic Regression Analysis, Residential Sector, Ethnicity of the household head, Education Level of the Household Head, Telephone facilities in the household area, Pipe borne line (main line) nearby household area, Any Household member engage to agricultural activity, Age of the Household Head and Household size are significantly effect on the probability of a poverty status of the household while gender of the household head, marital status of the household head, any of the household member receive income as an employee and household head suffer from chronic illness/disability are not statistically significant.

The results of the study concluded that probability of being poor households increases with the living in rural area, uneducated household head, not having telephone facilities and pipe borne line in the living area, ethnicity of the household head is not Sinhala, no one of the household member is engaged to agricultural activity, larger household size and younger household heads. Also this study is recommended that the Sri Lankan government should pay more attention on the education of the people, utility facilities of the general public.

Key words: Poverty, Logistic regression, Log likelihood, Odds ratio

Acknowledgements

I would like to acknowledge the efforts, support, guidance, cooperation and encouragement of numerous people who have made it possible for me to undertake this study.

I wish to express my sincere appreciation and gratitude firstly to my supervisor, Mr. Rohana Dissanayake, senior lecturer at department of Mathematics, University of Moratuwa for the valuable advices, guidance and suggestions given to me throughout the research study and also my course coordinator, Mr. T. M. J. A. Cooray for his valuable support in many ways.

I also would like to express my gratitude to the Mr. C.P. Ranathunga, Deputy Director and Mr. Mangala Ranathunga, Statistician at Department of Census and Statistics for their valuable support to take the data for this study. Besides all the staff at Data dissemination unit in Department of Census and Statistics who support to get the data needed for this research are also greatly appreciated.

Next, I expand my special gratitude to Ms. N.A.M.R. Senaviratna for the valuable support given to me in many ways to fulfill my task.

I would like to gratefully acknowledge the staff of Planning Unit, Central Environmental Authority specially Ms. D.C.N. Ranasingha, Ms. C.P.M. Perera and Ms. Hasanthi Walpita for their support given to me.

Finally, I would like to express my deepest gratitude to my parents, my husband and all other family members for their support and affection all long.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CI - Confidence Interval

CPH - Census of Population and Housing

DCS - Department of Census and Statistics

HIES - Household Income and Expenditure Survey

ML - Maximum likelihood

OPL - Official Poverty Line

OR - Odds Ratio

ROC - Receiver Operating Characteristic

SDG - Sustainable Development Goals

VIF - Variance Inflation Factor

LIST OF APPENDICES

CHAPTER 1
INTRODUCTION

1.1 Overview

This chapter presents the background of the study, statement of the problem, objectives and significance of the study. The organization of the report is presented at the end of the chapter.

1.2 Background of the Study

Poverty is a complex and multidimensional social phenomenon. Poverty is one of the biggest and most challenging problem and obstacle to human development. Effects of poverty are harmful to both individuals and society. “End poverty in all its forms everywhere by 2030” is the first goal in Sustainable Development Goals (SDG) adopted by all United Nations member states in 2015. This indicates that fighting poverty has become a global theme while the number of people living in extreme poverty decreased by more than half between the year 1990 and 2015 from 1900million to 836 million (UNDP, 2019).

Poverty status of the household is determined by comparing the monthly real expenditure of the household to the Official Poverty Line (OPL) multiply by number of members in the household. OPL is defined as the fixed real per capita expenditure per month at a specific welfare level with the expenditure for the consumption of food and non-food items. If the monthly real expenditure of the household is less than the value of the OPL multiply by number of members in the household, then that household is considered to be in poverty.

In Sri Lanka, 2.1% of the households were poor in the year 2016 and it was approximately 0.1694 million households. The poverty index is shown a downward trend from year 2002 to 2016. Household Income and Expenditure Survey (HIES) reveals that approximately 0.844 million individuals were in poverty in the year 2016 and 1.3 million in 2012/2013. This indicates 0.5 million (35.1%) decline from 2012/2013 to 2016.

Even though in the last few decades poverty has considerably dropped down at the national level, poverty disparities even now present at the provincial level and district level. Government has implemented some programs to mitigate poverty by providing financial and infrastructure support to those with low incomes. To build strategies aimed at poverty reduction and amenable to modification by policies, it is necessary to identify strongly associated factors with poverty.

1.3 Statement of the Problem

One of the greatest challenges of the humanity is eliminating poverty in all its forms. To address poverty, we have to identify the factors significantly related to poverty.

The government initiated some programs to increase consumption and self-employment among the poor and increased government expenditure on health and education. Old food stamps scheme is the first programme to address the poverty. Then this scheme was replaced by the Janasaviya Programme and credit facilities were provided for poorest people through the World Bank funded Janasaviya Trust. The government also tried to seek a solution to the development gap between urban and rural areas by providing bridge the development gap between the urban and rural areas by providing inducements for industries to locate in rural areas. With the change of government in 1995, Samurdhi program was introduced as a replacement of Janasaviya program. Samurdhi program is consisting of a small rural infrastructure component, a large income transfer component and a series of pro-poor credit schemes including the Grameen type Samurdhi Bank Scheme.

Government is trying to prevent poverty by giving financial support and infrastructure facilities for the poor household. But poverty disparities still exist across the provinces and districts. This indicates that the government should change the existing approaches to poverty prevention.

Though there is ample research on the poverty in Sri Lanka, most of the researches have done descriptive analysis only and some researchers have constructed the Poverty Index/Wealth Index using Principle Component Analysis. Also, most of the researches have considered only household income and expenditure to study the

poverty. But socioeconomic factors can be affected to household poverty. Therefore, this study statistically explored the significant socioeconomic and demographic factors which affect household poverty and attempt to fill the gaps by proposing solutions to the problem.

1.4 Objectives of the Study

The objectives of this study are as follows.

- Identify the socioeconomic and demographic factors that mainly associated with household poverty in Sri Lanka
- Estimate the effect of the statistically significant factors on household poverty

1.5 Significance of the Study

Poverty has many aspects and different factors include unemployment, social exclusions, natural disasters, diseases, characteristic of household head, demographic factors, etc. may be causes to the poverty. Therefore, identification of determinants of poverty is more productive to prevent the poverty in the society.

It is very significant if the studies of poverty are done for identifying the factors which significantly affect to the poverty. Because, it is enable to government and policymakers to modify regulations & laws and establishing preventive approaches & strategies to reduce poverty.

Determination of how poverty households are distributed in the country is very valuable things for Sri Lanka. Because increase of number of poor households is toughly affect to the country's economic growth. Therefore, it is very reasonable to study the determinants of poverty in Sri Lanka.

1.6 Organization of the Report

Chapter 2 describes the researches related to this study done in the past in Sri Lanka as well as in other countries.

Chapter 3 reviews the methodology used in the study. It reviews some of the theories and theoretical model based on the study. It also describes the data set and variables used in the study.

Chapter 4 explains the data analysis done to achieve the objectives. This chapter describes what factors are significantly effects on household poverty and how to find a suitable model to identify household wealth.

Chapter 5 describes the findings of the analysis of household poverty. It presents the significant factors associated with household poverty in Sri Lanka.

Finally, chapter 6 reports the conclusions and recommendations of the study.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

This chapter presents a literature review related to the statistical analysis of household poverty. This literature review investigated published research papers and Journal articles related to this study. In this literature review will be revealed about poverty in Sri Lanka and some interesting researches which are related to finding the behavior of poverty and identify the factors affecting poverty.

2.2 Poverty in Sri Lanka

Definition of poverty of a person is one who cannot spend a usual or socially acceptable minimal amount of money or materials in a particular period. Also, poverty exists when people cannot achieve their basic needs. To identify the poorest people, it is need to determine what constitutes basic needs. Constitutes basic needs may be defined as narrowly as "those necessary for survival", or as broadly as "those reflecting the prevailing standard of living in the community" (Augustyn, 2019).

The HIES - 2016 of the Department of Census and Statistics (DCS) reveals that the poverty rate in Sri Lanka in 2016 was 4.1 percent and it was declined from 6.7 percent in 2012/13. Even though in last few decades poverty has considerably dropped down at the national level, poverty disparities even now present at the provincial level and district level. This is explained by the Table 2.1.

Table 0.1: Information on poverty by country, residential sectors, provinces and districts in the year 2016

	Sector/ Province/ District	**Poverty Head Count Index (%)**	**Poor Population (Number)**	**Contribution to Total Poverty (%)**
Country	**Sri Lanka**	**4.1**	**843,913**	**100.0**
Residential Sector	Urban	1.9	67,649	8.0
	Rural	4.3	693,956	82.2
	Estate	8.8	82,308	9.8

Province	Western	1.7	101,342	12.0
	North Western	2.7	64,638	7.7
	Southern	3.0	74,769	8.9
	North Central	3.3	42,191	5.0
	Central	5.4	142,044	16.8
	Uva	6.5	83,885	9.9
	Sabaragamuwa	6.7	133,149	15.8
	Eastern	7.3	118,061	14.0
	Northern	7.7	83,834	9.9
District	Colombo	0.9	19,796	2.3
	Mannar	1.0	1,005	0.1
	Hambantota	1.2	7,450	0.9
	Gampaha	2.0	45,827	5.4
	Vavunia	2.0	3,526	0.4
	Puttalam	2.1	16,708	2.0
	Polonnaruwa	2.2	9,051	1.1
	Ampara	2.6	17,431	2.1
	Kalutara	2.9	35,719	4.2
	Galle	2.9	30,775	3.6
	Kurunegala	2.9	47,930	5.7
	Anuradhapura	3.8	33,140	3.9
	Matale	3.9	19,357	2.3
	Matara	4.4	36,544	4.3
	Kandy	5.5	76,429	9.1
	Moneragala	5.8	27,187	3.2
	Nuwara Eliya	6.3	46,257	5.5
	Ratnapura	6.5	72,715	8.6
	Badulla	6.8	56,698	6.7
	Kegalle	7.1	60,435	7.2
	Jaffna	7.7	46,052	5.5
	Trincomalee	10.0	39,718	4.7
	Batticaloa	11.3	60,912	7.2
	Mullaitivu	12.7	12,003	1.4
	Kilinochchi	18.2	21,249	2.5

Source: (Poverty Indicators: Household Income and Expenditure Survey - 2016, Department of Census and Statistics, 2017)

The OPL of Sri Lanka is Rs. 4,166 for the year 2016.According to the Table 2.1, 4.1% of the people in the country still lives below the OPL. Therefore, it needs to give attention to preventing poverty in Sri Lanka. To preventing poverty, it needs to identify the causes of poverty and government should establish preventive programs to address those causes.

2.3 Literature Review of Researches Related to Analysis of Poverty

2.3.1 Researches related to the poverty in Sri Lanka

To find the determinants of poverty in Sri Lanka and to quantify effects, (Deepawansa, Sooriyarachchi, & Wickremasinghe) used Binary logistic regression modeling, Principal Component Analysis and Factor Analysis. In this research, poor/non-poor household was used as response variable and household size, education level of the most educated person of the household, total income, per capita income, number of income earners and age of the household head were used as explanatory variables. The results of the binary logistic regression model show that, the most important determinant of being poor is household size and it is positively related with poverty status. The variables education level of the most educated person of the household, total income, per capita income, number of income earners and age of the household head are negatively related with poverty status. According to the results of factor analysis and principal component analysis, all considered variables extracted into clearly interpretable four factors such as household head characters, household factors, economic and number of income receivers for identifying the status of poverty.

Kumara & Gunewardena (2017) has studied the Disability and Poverty in Sri Lanka. In this research, they estimated multidimensional and monetary poverty among the household with disabled person by using 2006/07 and 2009/10 HIES data. The standard foster Greer Thorbecke indexes, namely poverty headcount, gap and severity indices were estimated in estimating monetary poverty. Using the standard Alkire Foster Approach, multidimensional poverty measures such as multidimensional poverty incidence, intensity and acute multidimensional poverty were calculated. This study concluded that households with disabled persons are

more tend to be poor than the household without disabled person and reduction of prevalence of monetary poverty among households without disabled person is higher than that of among households with disabled person. This was true for both HIES years at national level and also each residential sector.

Velnampy & Achchuthan (2013) have investigated the causes of poverty reduction between 2007 and 2010 in Sri Lanka. Covering 300 households in two districts in estate areas of Sri Lanka, a household survey was conducted by the author. This study was estimated poverty in Sri Lanka using multidimensional approach. The Alkire and Foster multidimensional poverty estimation method was used to examine the appropriate statistical technique to select the dimensions and weights of the variables and Polychoric Principle Component Analysis was selected. The results confirm that, there is a significant reduction of poverty level in estate areas between year 2007 and 2010. By comparing author's survey data with HIES 2006/07 and HIES 2009/10, it is suggested that poverty level in the estate areas are considerably decreased in recent years. This study was identified the major determinants of poverty in Sri Lanka. Those are education, access to household utilities, gender of the household head, household size, living area and province, receiving remittances and employment. Also, Logit and Probit analysis were used to identify the factors significantly related to the monetary poverty and multidimensional poverty. According to the results of Logit and Probit analysis, health related variables, gender of the household head, land availability and household size had opposite results for monetary poverty and multidimensional poverty.

To identify the micro level factors associated with household poverty in Sri Lanka, (Ranathunga, 2011) used Ordinary Least Square, quintile and probit regressions. Results of this analysis reviled human capital related factors are the major determinants of household poverty in Sri Lanka. Education level of the household head and education level of the other family members are negatively associated with household poverty in each residential sector. That is all the education related variables are statistically significant and positively related with improving living standard of the household in each residential sector. Larger household size and higher dependency ratio are tending to be poor and household with less than two

children are less like to be poor. Also, this study identified that the self-employment in each residential sector is negatively related with standard of living and gender of the household head is not a statistically significant determinant of the poverty in the urban sector.

Silva, 2008 have examined Micro-level determinants of poverty reduction in Sri Lanka. This study is used Sri Lanka integrated survey data conducted by the World Bank in 2000 to identify the poverty determinants. Using logistic regression analysis, this study revealed that household head education level, being engaged in business and being salaried employment are significantly affected to the poverty. Also, this study is identified that the probability of being poor household increases with the casual earner household head, female household head, living in rural area and household size.

Semasinghe, 2010 has used Qizilbash's "core poor" framework to determine the dimensions of rural poverty. Finding of this study revealed that the clean drinking water, food, cloths, health care, sanitation, housing, agricultural land, income and education & knowledge are the critical aspects of well-being of rural people.

2.3.2 Researches related to poverty in foreign countries

Joshi, Keshav Lall, &Luni (2012) was analyzed income and consumption measures of poverty by considering a case of Baitadi district from far western rural hills of Nepal. To identify the significantly associated variables with income and consumption poverty, binary logistic regression model was applied. Results of the study revealed that the livestock holding, operational landholding and household size are major determinants of food insecurity and dependency ratio and occupation are identified as major determinants of income poverty. Also, this study identified that landholding and education level of the household head are major determinants of both income and consumption poverty.

Logistic regression and linear log regression analysis were used to determine the most serious factors that influence poverty by Xhafaj & Nurja in the year 2014. As dependent variable, Expenditure of consumption of the household per capita used at

the linear log regression model and economic status(poor and non-poor) used at the logistic regression model. For both regression models, the demographic, educational, zone variables were used as independent variables. The results of this study revealed that the household size, educational level of the household head, gender of the household head and zone are most serious factors influence both Expenditure of consumption of the household per capita and the economic status.

Socioeconomic Determinants of Poverty in Paraguay has been studied by Duarte in the year 2016. In this study, researcher has analyzed the Paraguay's 2013 permanent survey of household data by using ordinary least square regression and logistic model. The results of this study indicated household head is female, working in agriculture, lacking a work contract, being self-employed, being only Guarani-speaking, lacking health insurance, lacking access to sewers and lacking education are the significant socioeconomic determinants of poverty.

Using 2003/2004 National Living Standard Survey data in Nigeria, Osowole, Uba, & Ugbechie in the year 2012 have been identified possible determinants of poverty. Logistic Regression Analysis was used for the study and the results reviled that the education level of the household head and household size are the most significant determinants of poverty. Other significant determinants of poverty are father's education level, mother's occupation, father's occupation, gender of the household head, age of the household head and occupation group of the household head. This study recommended that get formal education and keep moderate household size to reduce poverty.

Mok, Gan, & Sanyal (2007) has used logistic regression analysis to identify the determinants of urban poverty in Malaysia. A sample of 2,403 urban households was selected from the Household Expenditure survey: 2004-2005 for this study. Findings of this study showed that migrant workers are more chance to be poor and human capital significantly reduce the probability of being poor. Other important determinants of urban poverty in Malaysia are region, race and household size.

There are many studies of poverty determinants in foreign countries ((Geda, Jong, Mwabu, & Kimenyi, 2001), (Achia, Wangombe, & Khadioli, 2010), (Hashmi,

Hashmi, & Sial, 2008), (Dudek & Lisicka, 2013) etc.). Most of the poverty-based studies were used logit regression methods as statistical technique to identify the determinants of poverty at the household level. Also, most commonly used dependent variable is binary indicators of poverty status and most of the studies are reviled that poverty status is highly associated with household size, education level of the household head, age of the household head, occupation of the household head, ethnicity and engagement in agricultural activity.

2.4 Summary

Researchers in different part of the world conducted many numbers of researches with poverty but few researches in Sri Lanka. A careful review of these studies had enabled me to gain some important insights relevant to my research.

Reduction of poverty is one of the major objectives of development. Therefore, it is necessity to understand and analyze the current status of poverty in a country. Also, it is necessary to identify the determinants of poverty to take prominent action against the poverty.

According to past research, I could identify some factors which affecting to the status of poverty. Such as household size, education level of the household head, gender of the household head, age of the household head, occupation group of the household head, Ethnicity and engagement in agricultural activity.

Logistic regression model was used by most of the researchers to identify the determinants of poverty. Therefore, it is confirmed that Logistic regression model is a most suitable statistical technique to identify the significant factors which affect to the poverty in a particular country.

CHAPTER 3
METHODOLOGY

3.1 Overview

This chapter presents the source of data, data description, methods carried out for research, statistical tests carried out in the analysis.

3.2 Source of Data

Secondary data were used in this study to accomplish the objectives and those data were acquired from HIES-2016 conducted by DCS in Sri Lanka. The HIES - 2016 is the 9^{th} survey in the HIES series. Data collection part of the survey was carried out from January to December, 2016 covering all 25 districts with 21,756 households in the country. Two stage stratified sample design is used and residential sectors in each district of the country are the selection domain for the stratification. The sampling frame of this survey was the list of households prepared for the Census of Population and Housing (CPH), 2011.

3.3 Data Description

Response variable of this study is status of the household poverty consists of two levels namely poverty and non-poverty. These two levels were calculated using the OPL for 2016 in Sri Lanka. OPL is the real per capita expenditure per month for a person fixed at a specific welfare level with the consumption expenditure of food and non-food items. The OPL is Rs.4,166 for the year 2016 in Sri Lanka (Poverty Indicators: Household Income and Expenditure Survey - 2016, Department of Census and Statistics, 2017). If expenditure of a household is less than or equal the OPL multiplied by household size is named as poverty and if expenditure of a household is more than the OPL multiplied by household size is named as non-poverty.

Sample size of the household in the HIES-2016 is 21,756. Data was collected from a questionnaire and questionnaire divided in to 9 sections as follows.

Section 1: Demographic Characteristics

Section 2: School Education (For people aged 5-20 years)

Section 3: Health

Section 4: Expenditure

Section 5: Income

Section 6: A - Inventory of durable goods

B - Indebtedness of the Household

Section 7: Access to Primary facilities

Section 8: Housing Information

Section 9: Agricultural Holdings & Livestock

In section 5, information on income of the household is taken by using 6 categories. Basically, it is performed descriptive statistics and graphical analysis roughly on income of the household. It can be identified that 82 households are not given their income in any of the categories at section 5 and only one poor household is in this 82 households. Income is most related to the poverty. Therefore, if we include these 82 household for this analysis it will mislead the results. Therefore these 82 households are removed from the dataset for the study.

Identify the socioeconomic and demographic factors that significantly affect household poverty in Sri Lanka is the main objective of the study. Thus, to accomplish the objective following factors are considered.

Table 0.1: Description of the variables

No.	**Factor**	**Levels**	**Abbreviation**
01	Residential Sector	Urban	Ur
		Rural	Ru
		Estate	Es
02	Gender of the household head (Gender)	Male	M
		Female	F
03	Ethnicity of the household (Ethnicity)	Sinhala	Sin
		Tamil	Tam

		Sri Lankan Moors	S_MO
		Malay	Ma
		Burgher	Bu
		Other	Oth
04	Marital status of the household head (MS)	Never Married	NM
		Married	Mari
		Widowed	Wid
		Divorced	Diw
		Separated	Sep
05	Education level of the household head (Edu_Lv)	No Schooling	NS
		Primary Education	PE
		Up to O/L	OL
		Up to A/L	AL
		Higher Education	HE
06	Main activity of the household head (M_Act)	Engaged in economic activity	EA
		Seeking for and available to work	SW
		Household activities	HA
		Retired	Re
		Unable to work (Too old / Disable)	UW
		Other	Oth_Ac
07	Any of the household member engage to agricultural activity (EAgri)	Engage	EAgri_Y
		Not engage	EAgri_N
08	Any of the household member receive income as employee (HMInc)	Yes	HMInc_Y
		No	HMInc_N
09	Household head suffer from	Yes	CI_Y

	chronic illness/Disability (CI)	No	CI_N
10	Child death in the household (CD)	Yes	Ch_D
		No	Ch_ND
11	Electricity supply near the area (ES)	Yes	ES_Y
		No	ES_N
12	Telephone facilities in the area (TelF)	Yes	TelF_Y
		No	TelF_N
13	Pipe borne line near by the area (WatF)	Yes	WatF_Y
		No	WatF_N
14	Age of household head (Age)		
15	Household size (HZ)		
16	Household Income (HHI)		

Age, Household size and Household income are continuous variables among the variables considered in this study. Dummy variables are used to represent the categorical variables in the analysis.

3.4 Research Methodology

To identify the socioeconomic and demographic factors that mainly associated with household poverty in Sri Lanka, several statistical techniques were used throughout this study. Data analyses are arrayed mainly under descriptive analysis and advance analysis.

In descriptive analysis, bar charts, frequency tables and cross tabulations were used to represent the data.

Due to the dichotomous nature of the dependent variable, Binary Logistic Regression Analysis is carried out as advanced analysis to investigate the combined effect of the independent variables on status of the household poverty. Before performing the Binary Logistic Regression Analysis, One Sample Proportion Test is used to reduce the levels of factors that influence the poverty for easy interpretation of the binary

logistic regression model and univariate analysis techniques such as Mann-Whitney U Test & Chi-square Test are performed to check the association between each contributory factor and the status of the household poverty.

The statistical software, MINITAB version 18.0 and SPSS version 20.0 were used for the analysis of data.

Descriptive analyses and one sample proportion test were performed using MINITAB software. SPSS software was used to perform Mann-Whitney U Test, Chi-square Test and binary logistic regression analysis.

3.5 Brief Description of Statistical Techniques Used in the Study

3.5.1 Binary logistic regression

Binary Logistic Regression analyzes the relationship between dichotomous dependent variable and multiple independent variables while estimating the probability of occurrence of an event by fitting data to a logistic curve. The independent variables are either continuous or categorical and dependent variable is the population proportion/probability that the resulting outcome is equal to 1. Parameters obtained for the independent variables can be used to estimate odds ratios for each of the independent variables in the model.

The specific form of the logistic regression model is:

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots.\beta_n x_n}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots.\beta_n x_n}}$$

Where; π - Probability of the outcome of an event

β_0 - Intercept

$\beta_1, \ldots, \beta_n$ - Regression Coefficients

$x_1, x_2, \ldots, x_n$ - Independent Variables

The transformation of the conditional mean $\pi(x)$ logistic function is known as Logit Transformation.

$$\ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n$$

The importance of the logit transformation is that it is linear in its parameters, and may range from $-\infty$ to $+\infty$.

Assumptions of Binary Logistic Regression:

There are four assumptions under binary logistic regression. These assumptions are required to satisfy to give a valid result.

- **Linearity:** Any of the independent variables should have linear relationship with the logit of the dependent variable. If the relationship between log odds of the dependent variable and any of the independent variables is not linear, the model will not be accurate
- **Independent errors:** Errors should not be correlated for each pair of observations
- **Multicollinearity:** Independent variables should not be highly correlated with each other
- **No Outliers:** There should not outliers, high leverage values or highly influential points.

Following terms and tests were used when perform the binary logistic regression analysis.

a) Detecting of Multicollinearity

Multicollinearity is a statistical phenomenon in which two or more independent variables in a multiple logistic regression model are highly correlated or associated. (Midi, Sarkar, & Rana, 2013). Multicollinearity can be detected with tolerance and it is reciprocal defined as Variance Inflation Factor (VIF).

Tolerance of any specific independent variable is "1-R^2". R^2 is the coefficient of determinant for the regression of that independent variable on all other remaining

variables. Tolerance close to 1 indicates that there is little multicollinearity, whereas a value close to zero suggests that multicollinearity may be a threat. (Midi, Sarkar, & Rana, 2013).

$$VIF = \frac{1}{TOLERANCE}$$

VIF shows that how much the variance of the coefficient is being inflated by multicollinearity. There is no formal cutoff value to use with VIF for determining the presence of multicollinearity. Values of VIF exceeding 10 are often regarded as indicating multicollinearity, but in weaker models, which is often the case in logistic regression; values above 2.5 may be a cause for concern (Midi, Sarkar, & Rana, 2013).

b) Variable Selection Methods

Forward selection and backward elimination are the commonly used variable selection methods in logistic regression. However, backward elimination is often less successful than forward selection because the full model fit in the first step is the model most likely to result in a complete separation of response values. Therefore, forward selection method is used in this study.

In forward selection, the score chi-square statistic is computed for each effect on the model and largest significant effect is added to the model and never removed from the model. This process is repeated until none of the remaining effect significant.

c) Odd and Odds Ratio

Odds of an event is the ratio of the probability that an event will occur to the probability that it will not occur. If π is the probability of an event occurring, $(1- \pi)$ is the probability of the event not occurring. Then corresponding odds value is given by;

$$\text{Odds of event} = \frac{\pi}{1-\pi}$$

The effect of independent variables on dependent variable is explained in terms of odds, because the logistic regression calculates the probability of an event occurring over the probability of the event not occurring.

The Odds ratio (OR) is a measure of association between exposure and outcome. In the logistic regression, the exponential function of the regression coefficient (e^{β_1}) is the OR associated with a one unit increase in the independent variable when other independent variables are held constant. It indicates the relative amount by which the odds of the outcome increase (OR>1) of decrease (OR<1) when the value of the corresponding independent variable increase/decrease by 1 unit.

d) Measures of Goodness of Fit

To measure the goodness of fit of the model, Chi-square goodness of fit test, Likelihood ratio test and Hosmer and lemeshow test were used in this study.

i) Chi-square goodness of fit test

In logistic regression, the deviance is used as the statistic for overall fit of the model instead of R^2 and the chi-square is used as a measure of goodness of fit of the observed values to the expected values. The model is lack of fit, when the difference between observed value and expected value (deviance) is large. Therefore, small deviance is expected for good fit of the model. When more variables are added to the model the deviance will get smaller.

ii) Likelihood ratio test

A logistic regression model with the k independent variables is selected to provide a better fit to the data if it demonstrates an improvement over the model with the null model. The overall fit of the model with k coefficients can be examined using likelihood ratio test which tests the null hypothesis

H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$.

The deviance of the null model is compared with the deviance of the model with k independent variables to test the above hypothesis. To measure effect of k

independent variables to the dependent variable, Goodness of fit index (G) is measured.

G = (-2 log likelihood of null model) - (-2 log likelihood of model with k variables)

iii) Hosmer and lemeshow test

The Hosmer and Lemeshow test is used to examine whether the observed proportions of events are similar to the predicted proportions of occurrence in subgroups of the model population. It indicates the extent to which the estimated model provides a better fit to the data (i.e. better predictive power) than the null model. The test statistic asymptotically follows a χ^2 distribution (Hosmer, Hosmer, Lemeshow, & Cessie, 1997).

The formula of the test statistics is

$$H = \sum_{g=1}^{10} \frac{(O_g - E_g)^2}{E_g}$$

Where O_g and E_g denote the observed event and expected event for the g^{th} risk decile group.

e) Model summary statistics

Following are the model summary statistics used in this study.

i) Cox-Snell R^2

The Cox and Snell R^2 is a pseudo R^2 statistic 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. It has the limitation that it cannot achieve the value of 1.0 as R^2 in linear regression.

It is given by:

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

Where,

L(R) – Likelihood of the model has only intercept

L(F) – Likelihood of specified model

N – Number of observations

ii) Nagelkerke R^2

The Nagelkerke R^2 adjusts the Cox-Snell R^2 so the range of possible values extends to one.

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

Where,

L(R) – Likelihood of the model has only intercept

L(F) – Likelihood of specified model

N – Number of observations

f) Receiver Operating Characteristic (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. The plot of sensitivity vs. (1-specificity) is called the ROC curve. The area under this curve provides an overall measure of fit of the model. It ranges from 0.5 and 1.0 with larger values indicative of better fit. The area under the ROC curve, which ranges from zero to one, provides a measure of the model's ability to discriminate the larger the area under the ROC curve, the more the model discriminates.

g) Model Validation of the Logistic Regression

When logistic regression analysis is performed using sample data and the developed model can be extended to the relevant population, the developed model is good fit. This is referred as model validation.

When a 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). In this study, data-splitting technique is used to validate the fitted model. In this technique, observations are randomly split in to two parts; one for develop the model and other one for measure the validation of the developed model.

h) Identifying influential observations

In logistic regression, Leverage values, Residual Deviance and change in parameter estimates are mainly used as a measure for identifying influential observations. The basic idea behind of this measure is to delete the observations one at a time, each time refitting the logistic regression model on the remaining n–1 observations. Then, the results are compared using all n observations to the results with the i^{th} observation deleted to see how much influence the observation has on the analysis.

3.5.2 One Sample Proportion Test

Some levels of factors that are influencing the household poverty may be neglected because of their small proportions. Therefore, one sample proportion test was used to reduce the levels of factors. The following hypothesis was tested:

$H_0: p_i = 0$ Vs. $H_1: p_i \neq 0$

where p_i is the proportion of level i within the factor.

3.5.3 Chi-Square Test

Pearson Chi-Square test was used to check the association between each independent variable and the response variable.

The formula of the Pearson Chi-square test statistic is

$$\chi^2 = \frac{(O_i - E_i)^2}{E_i}$$

Where;

O_i – Observed value

E_i– Expected value

3.5.4 Mann-Whitney U Test

Mann-Whitney U Test is a two independent sample nonparametric test that used for test whether there is a relationship between two independent samples/variables. When two independent variables are not normally distributed or the sample sizes are small and both variables either ordinal or continuous, this test is appropriate. This test compares the medians between two variables to test the following hypothesis.

H_0: There is no relationship between two independent samples/variables

H_1: There is a relationship between two independent samples/variables

CHAPTER 4
DATA ANALYSIS

4.1 Overview

This chapter presents the results obtained by analyzing status of poverty in Sri Lanka. The first part of this chapter presents the results of descriptive analysis and the second part discusses the results of significant factors affecting on poverty status of the household, goodness of fit measures, model diagnostics and predictive accuracy.

4.2 Descriptive Analysis

Before carrying out the advanced analysis it is important to get a better idea about the background of the dataset.

4.2.1 Composition of the sample

The sample of the HIES is composed of household data obtained from 25 districts in Sri Lanka in the year 2016. There are 21756 households have covered in this HIES. According to the table 4.1, 2.1% households are poor in Sri Lanka in 2016.

Table 0.1: Composition of the sample of HIES

Poverty status of the Household	Frequency	Percentage
Non-Poverty	21301	97.9
Poverty	455	2.1
Total	21756	100.0

As mentioned in section 3.3, 82 households were removed from the dataset due to wrong information on income. Table 4.2 illustrate that the composition of the dataset used for this study.

Table 0.2: Composition of the sample taken for the study

Poverty status of the Household	Frequency	Percentage
Non-Poverty	21220	97.9
Poverty	454	2.1
Total	21674	100.0

According to the table 4.3, secondly highest non poverty households are in urban sector but in poverty, that is in Estate sector. Therefore, this result indicates that the residential sector is affect to the poverty status of the household. According to the table 4.3, most of the data are collected from rural sector as population of the rural sector is larger than other two residential sectors

Table 0.3: Classification of Poverty status by residential sector

			Residential Sector			**Total**
			Urban	**Rural**	**Estate**	
Poverty status	**Non-Poverty**	Count	3400	17006	895	21301
		% within Poverty status	16.0%	79.8%	4.2%	100%
		% within Sector	99.2%	97.8%	95.9%	97.9%
		% of Total	15.6%	78.2%	4.1%	97.9%
	Poverty	Count	29	388	38	455
		% within Poverty status	6.4%	85.3%	8.4%	100.0%
		% within Sector	0.8%	2.2%	4.1%	2.1%
		% of Total	0.1%	1.8%	0.2%	2.1%
Total		Count	3429	17394	933	21756
		% within Poverty status	15.8%	80.0%	4.3%	100.0%
		% within Sector	100.0%	100.0%	100.0%	100.0%
		% of Total	15.8%	80.0%	4.3%	100.0%

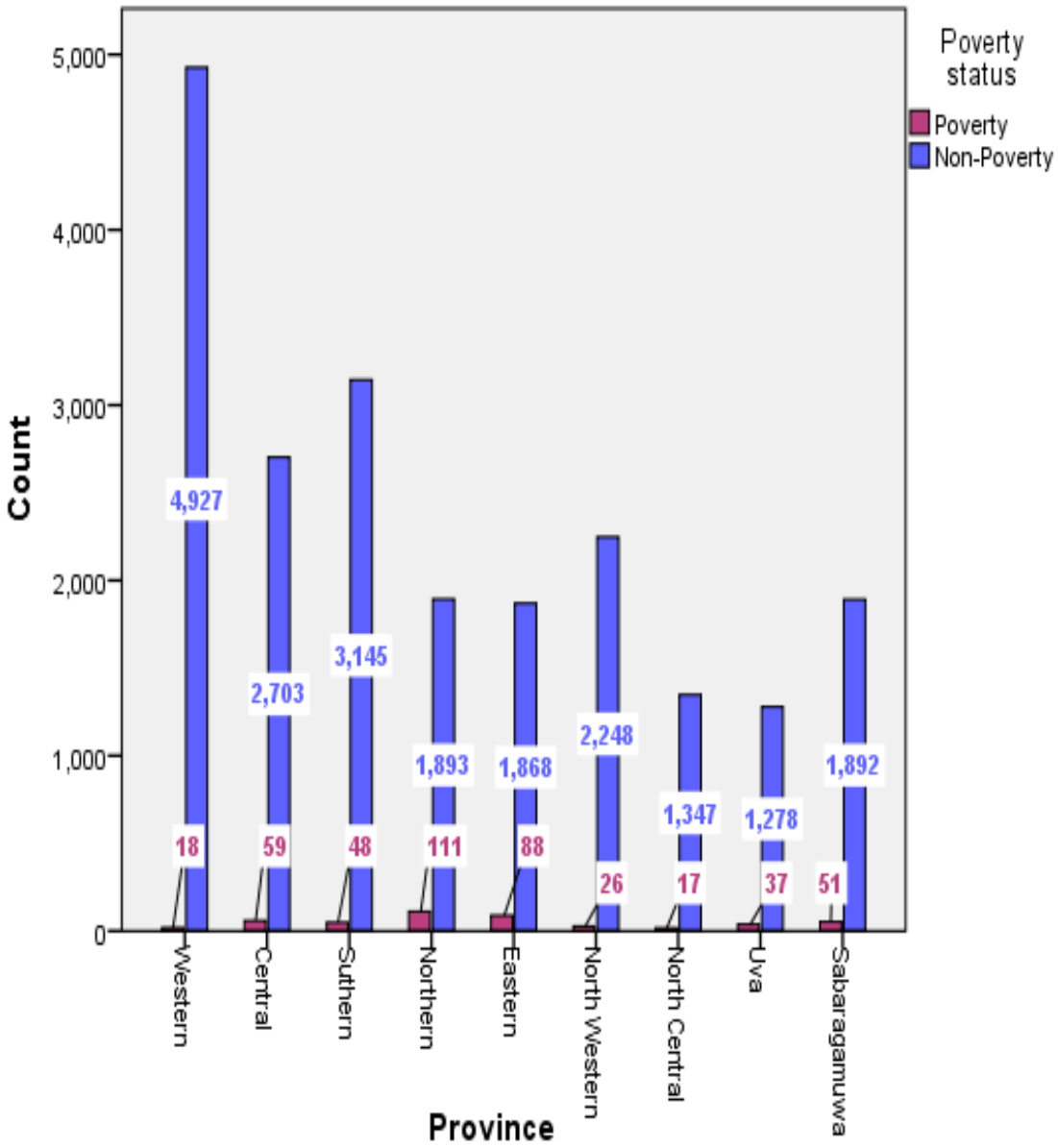

Figure 0.1: Classification of poverty status by province

According to the figure 4.1, out of the total poverty households, most of the poverty households are in Northern (24.2%) and Eastern (19.4%) provinces. (Related Cross tabulation is in Appendix A)

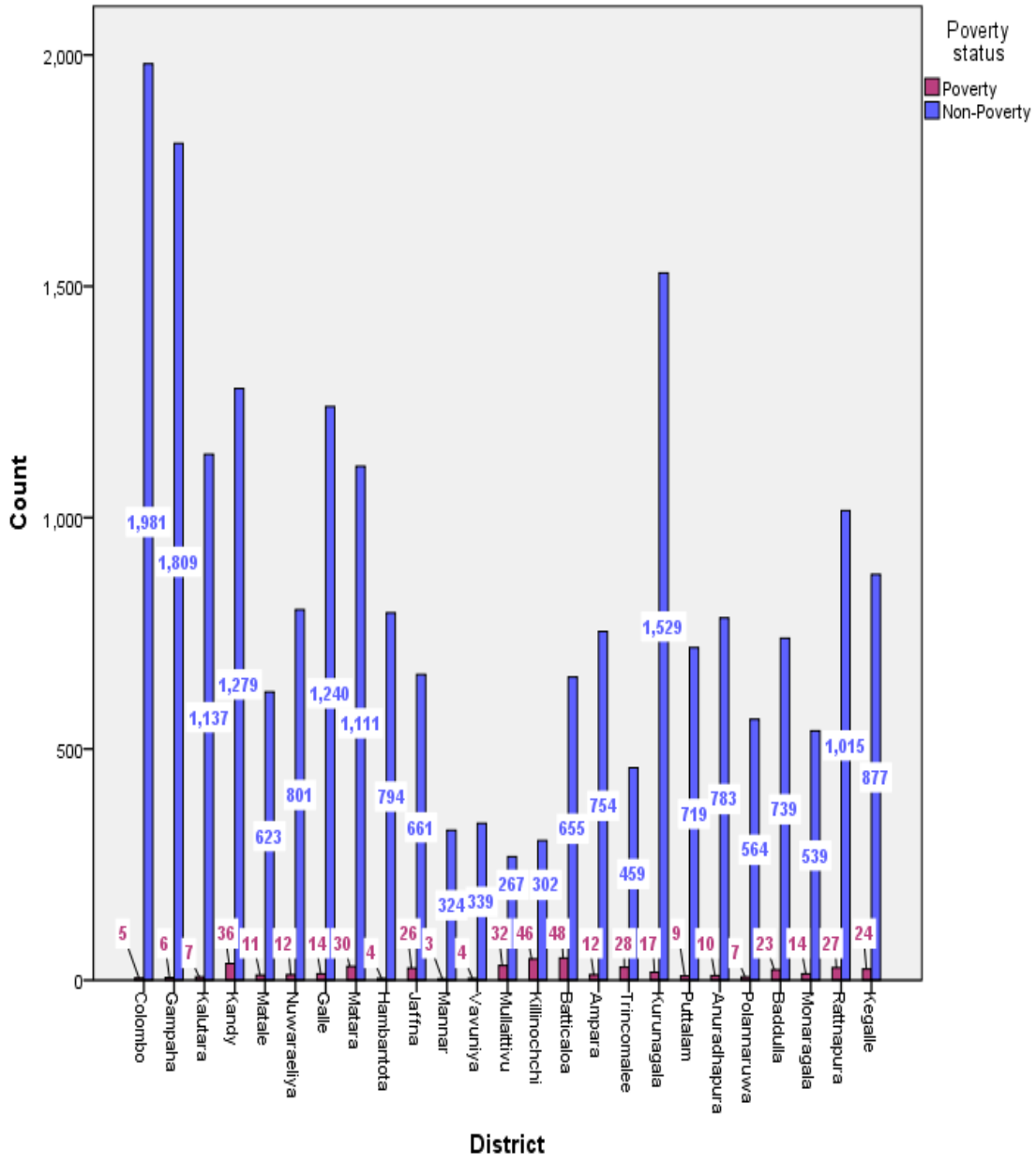

Figure 0.2: Classification of poverty status by District

Figure 4.1 concluded that the most of the poor households are in Northern. By considering figure 4.2, it can be identified that in Northern Province most of the poor households are in Kilinochchi. Secondly highest poor households are in Eastern province and in the Eastern province Batticaloa district have most poor household than the other district in Eastern. But out of the total district, most poor households are presented in Batticaloa district (10.5%). (Related Cross tabulation is in Appendix B)

Table 0.4: Classification of Poverty status by education level of the household head

			Education level of the household head				Total
			Primary Education	Up to O/L	Up to A/L	Higher Education	
Poverty status	Non-Poverty	Count	5467	12146	3051	637	21301
		% within Poverty status	25.7%	57.0%	14.3%	3.0%	100%
		% within Education level	95.7%	98.4%	99.9%	99.8%	97.9%
		% of Total	25.1%	55.8%	14.0%	2.9%	97.9%
	Poverty	Count	248	202	4	1	455
		% within Poverty status	54.5%	44.4%	0.9%	0.2%	100.0%
		% within Education level	4.3%	1.6%	0.1%	0.2%	2.1%
		% of Total	1.1%	0.9%	0.0%	0.0%	2.1%
Total		Count	5715	12348	3055	638	21756
		% within Poverty status	26.3%	56.8%	14.0%	2.9%	100.0%
		% within Education level	100.0%	100.0%	100.0%	100.0%	100.0%
		% of Total	26.3%	80.0%	14.0%	2.9%	100.0%

According to the table 4.4, when we consider non-poverty households, largest number of households are belonging to the household head has above primary to O/L education level but considering poverty households, largest number of households are belonging to the household head has below primary education. This result indicates that the education level is affect to the poverty status of the household. Also, it is concluded that when household head has below the primary education level that household is tend to be poor than other households (54.5%).

4.3 Advanced Analysis

The objective of this analysis is to identify the socioeconomic and demographic factors that mainly associated with household poverty in Sri Lanka. Poverty status of the household has 2 levels namely Non-poor and Poor. So, this variable is dichotomous variable and can't use linear regression model for achieve the objective. Binary logistic regression is more suitable for this study.

4.3.1 Reduction of factor levels

It is better to have few levels of factors for easy interpretation in logistic regression model. Therefore, one sample proportion test (Test Alternative hypothesis is proportion less than 0.1) is used to reduce the levels of factors having more than 2 levels and the test results are shown in the Table 4.4.

Table 0.5: Test results of the one sample proportion test

Factor	Levels	Frequency	Proportion	P-Value
Sector	Urban	3419	0.157747	1.000
	Rural	17325	0.799345	1.000
	Estate*	**930**	**0.042909**	**0.000**
Ethnicity	Sinhala	15718	0.725201	1.000
	Sri Lankan Tamil	3264	0.150595	1.000
	Indian Tamil*	**788**	**0.036357**	**0.000**
	Sri Lankan Moors*	**1816**	**0.083787**	**0.000**
	Malay *	**48**	**0.002215**	**0.000**
	Burgher *	**29**	**0.001338**	**0.000**
	Other *	**11**	**0.000508**	**0.000**
Marital Status	**Never Married ***	**477**	**0.022008**	**0.000**
	Married	16848	0.777337	1.000
	Widowed	3627	0.167343	1.000
	Divorced *	**142**	**0.006552**	**0.000**
	Separated *	**580**	**0.026760**	**0.000**
Education level	**No Schooling***	**833**	**0.038433**	**0.000**
	Primary Education	4856	0.224047	1.000
	Up to O/L	12297	0.567362	1.000
	Up to A/L	3051	0.140768	1.000
	Higher Education	**637**	**0.029390**	**0.000**

Main activity of the household head	Engaged in economic activity	15121	0.697656	1.000
	Seeking for and available to work*	**133**	**0.006136**	**0.000**
	Household activities	2587	0.119360	1.000
	Retired *	**1113**	**0.051352**	**0.000**
	Unable to work (Too old / Disable)	2476	0.114238	1.000
	Other *	**243**	**0.011212**	**0.000**

Note: Test Alternative hypothesis is proportion less than 0.1

* Significant at the 0.05 level of significance

Based on the results of Table 4.5, some factor levels can be neglected or merged because of their small proportions. Following table (Table 4.5) illustrates that how reduce factors and what are the new factor levels.

Table 0.6: Description of new factor levels

Factor	**Previous Levels**	**New levels**	**Frequency**	**Proportion**
Residential Sector	Urban	Urban	3419	0.157747
	Rural	Rural	17325	0.799345
	Estate	Estate	930	0.042909
Ethnicity	Sinhala	Sinhala	15718	0.725201
	Sri Lankan Tamil	Other Ethnicities (Othr_Eth)	5956	0.274799
	Indian Tamil			
	Sri Lankan Moors			
	Malay			
	Burgher			
	Other			
Marital Status	Married	Married	16848	0.777337
	Never Married	Single	4826	0.222663
	Widowed			
	Divorced			
	Separated			
Education level	**No Schooling**	Up to Primary Education (PE)	5689	0.262480
	Primary Education			
	Up to O/L	Up to O/L (OL)	12297	0.567362
	Up to A/L	Up to A/L (AL)	3051	0.140768
	Higher Education	Higher Education (HE)	637	0.029390

Main activity of the household head	Engaged in economic activity	Engaged in economic activity	15121	0.697656
	Seeking for and available to work	Not Engaged in economic activity	6552	0.302298
	Household activities			
	Retired			
	Unable to work (Too old / Disable)			
	Other			

4.3.2 Identification of the association between independent variables and dependent variable

There are three (3) continuous variables and thirteen (13) categorical variables considered as independent variables in this study. According to the Kolmogorov-Smirnov test (Appendix C), all continuous variables are not normally distributed. Therefore, "Mann-Whitney U test" is performed to test the relationship between continuous independent variables & dependent variable and "Chi-square test" is performed to test the relationship between categorical independent variables & dependent variable. According to these analyses, following table describes the relationship between each factor and the poverty status of households.

Hypothesis for Mann-Whitney U test

H_0: There is no association between continuous independent variable and poverty status of households

H_1: There is an association between continuous independent variable and poverty status of households

Hypothesis for Chi-square test

H_0: There is no association between categorical independent variable and poverty status of households

H_1: There is an association between categorical independent variable and poverty status of households

Table 0.7: Identification of the association between independent variables and dependent variable

Independent Variables	**Test Statistic**	**P value**
Residential Sector	χ^2 (2) = 45.520	0.000
Gender	χ^2 (1) = 4.312	0.038
Ethnicity	χ^2 (1) = 188.567	0.000
Marital Status	χ^2 (1) = 11.768	0.001
Education level	χ^2 (4) = 221.339	0.000
Main activity of the household head *	**χ^2 (1) = 0.054**	**0.816**
Engage to agricultural activity	χ^2 (1) = 5.806	0.016
Any of the household member receive income as an employee	χ^2 (1) = 20.681	0.000
Household head suffer from chronic illness/Disability	χ^2 (1) = 5.468	0.019
Child death in the household *	**χ^2 (1) = 0.389**	**0.533**
Electricity supply (main line) nearby household area*	**χ^2 (1) =3.193**	**0.074**
Telephone facilities in household area	χ^2 (1) =78.825	0.000
Pipe borne line (main line) nearby household area	χ^2 (1) =56.686	0.000
Household size	Mann-Whitney U test statistic=2,742,991.50	0.000
Age of the household head	Mann-Whitney U test statistic = 4,383,835.50	0.000
Income of the household	Mann-Whitney U test statistic = 3,162,133.00	0.000

Note: *Non significant at the 0.05 level of significance

According to the results of Table 4.6, it can be identified that Residential Sector, Gender, Ethnicity, Marital Status, Education level, Engage to agricultural activity, Any of the household member receive income as an employee, Suffer from chronic illness/Disability, Telephone facilities in household area, Household size, Age of the household head, Income of the household and Pipe borne line (main line) nearby

household area are significantly associated with the poverty status of the household. Only three factors namely Main activity of the household head, Child death in the household and Electricity supply (main line) nearby household area are not significantly correlated with the poverty status. Therefore, non-significant factors are removed and continued the analysis.

4.3.3 Detecting multicollinearity in binary logistic regression

One of the assumptions in logistic regression is that independent variables should not be highly correlated with each other. Therefore, before applying logistic regression, multicollinearity should be checked among explanatory variables. To check the multicollinearity, tolerance and VIF values are measured and Table 4.7 shows those results.

Table 0.8: Diagnostics tests results for selected models

Factor	**Levels**	**Dummy variable**	**Collinearity Statistics**	
			Tolerance	**VIF**
Sector	Ur Ru Es	S_Ur=1, Ur =0, Other	0.885	1.130
		S_Es=1, Es =0, Other	0.883	1.132
Gender	Male Female	Gen_M = 1, Male = 0, Female	0.580	1.724
MS	Single Married	MS_Sing = 1, Single = 0, Married	0.553	1.809
Ethnicity	Sin Oth_Eth	Eth_Tam = 1, Sin = 0, Oth_Eth	0.798	1.254
Edu_Lv	PE OL AL HEdu	EL_OL = 1, OL = 0, Other	0.852	1.173
		EL_AL = 1, AL = 0, Other	0.975	1.025
		EL_HEdu = 1, HEdu = 0, Other	0.991	1.009
HHSCI	CI_Y CI_N	HHSCI_Y=1, CI_Y =0, CI_N	0.873	1.146
TelF	TelF_Y TelF_N	TelF_Y = 1, TelF_Y = 0, TelF_N	0.828	1.208

WatF	WatF_Y WatF_N	Wat _Y=1, WatF_Y = 0, WatF_N	0.815	1.227
EAgri	EAgri_Y EAgri_N	EAgri_Y=1, EAgri_Y = 0, EAgri_N	0.834	1.199
HMInc	HMInc_Y HMInc_N	HMInc_Y = 1, HMInc_Y =0, HMInc_N	0.839	1.192
Age			0.705	1.418
HHZ			0.779	1.284
HHI			0.795	1.258

In this table we observe the all tolerance values are closer to 1 and all variance inflation factors are smaller than two (2). Therefore it is concluded that the data are free with multicollinearity.

4.3.4 Checking linearity assumption

One of the assumptions in logistic regression is continuous independent variables should have a linear relationship with the logit of the dependent variable. Household size, Age of the household head and income of the household are the continuous variable in this study. Therefore, the linear relationship between continuous independent variable and the logit of the dependent variable is checked by scatter plot.

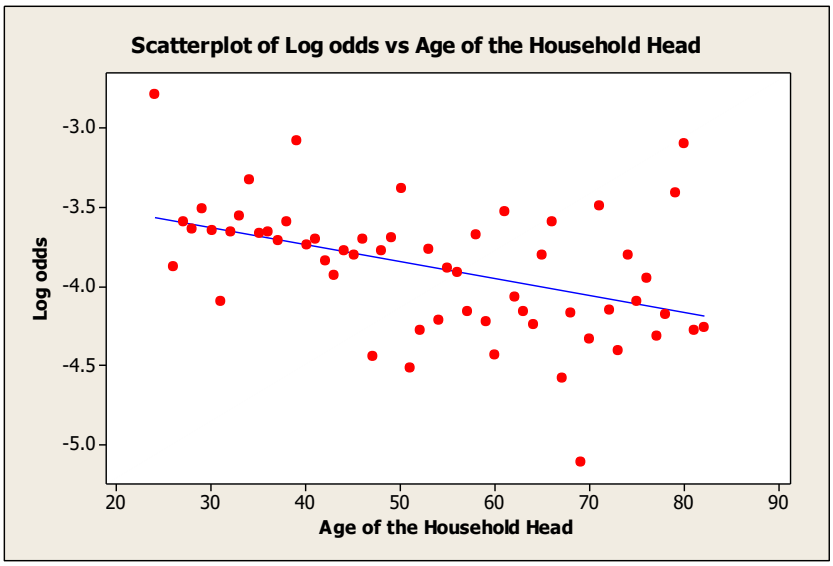

Figure 0.3: Log odds of poverty status Vs age of the household head

Scatter plot of log odds Vs age of the household head is shown in figure 4.3. It shows that variable age is quite linearly associated with the poverty status in logit scale. Thus, variable age can be added to the logistic regression model.

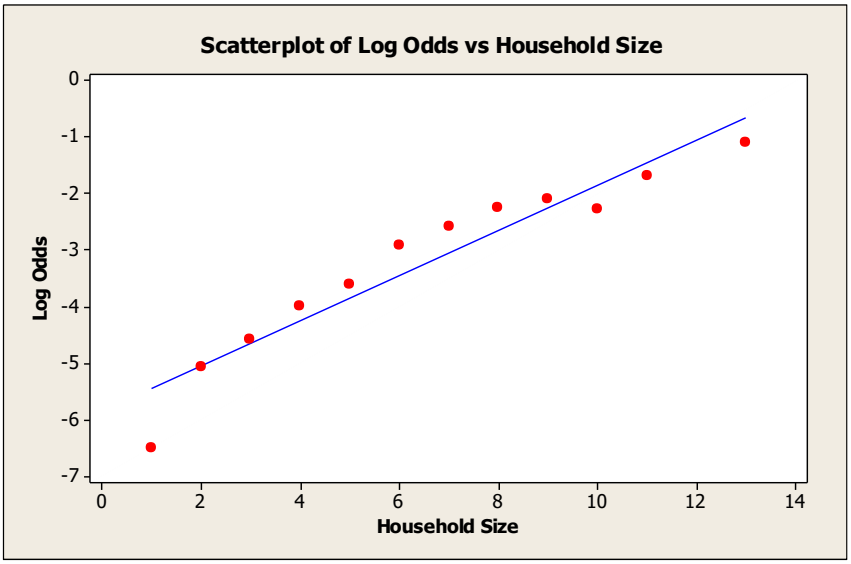

Figure 0.4: Log odds of poverty status Vs household size

Scatter plot of log odds Vs household size is shown in figure 4.4. It shows that variable household size is quite linearly associated with the poverty status in logit scale. Thus, variable household size can be added to the logistic regression model.

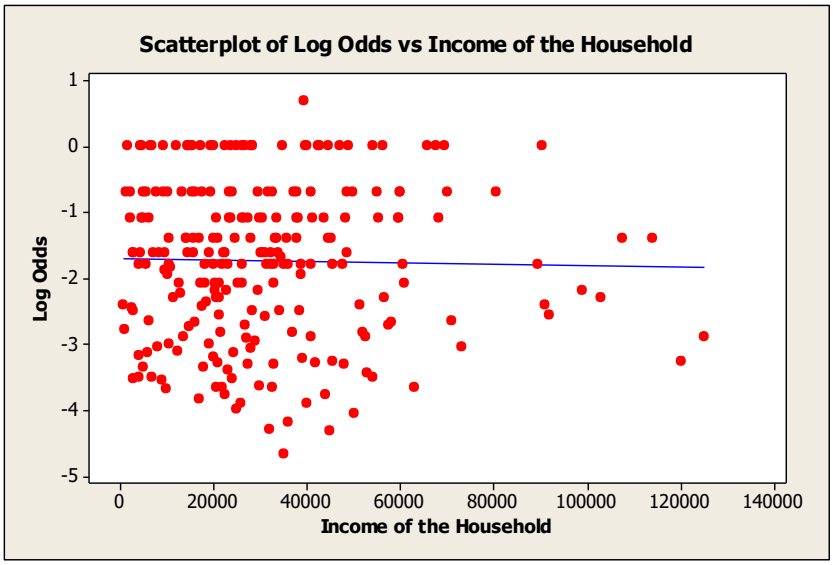

Figure 0.5: Log odds of poverty status Vs income of the household

Scatter plot of log odds Vs income of the household is shown in figure 4.5. It shows that variable income of the household is not linearly associated with the poverty status in logit scale. Thus, variable income of the household should be transformed.

Income of the household is transformed using box-cox transformation. Transformation variable of the household income is $(HHI)^{0.06}$. This variable is known as THHI.

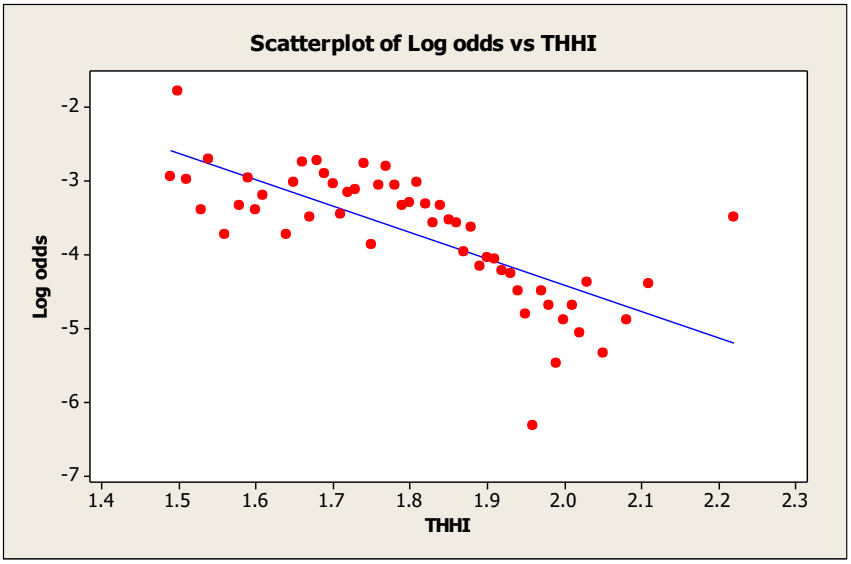

Figure 0.6: Log odds of poverty status Vs THHI

Scatter plot of log odds Vs THHI is shown in figure 4.6. It shows that variable THHI is linearly associated with the poverty status in logit scale. Thus, transformation of HHI can be added to the logistic regression model.

4.3.5 Binary logistic regression model

Since the dependent variable is dichotomous (Poverty and Non-poverty), the Binary Logistic Regression Model is used to fit the data. To estimate the parameters of the logistic regression model, maximum likelihood procedure is used. Forward stepwise selection method (Likelihood ratio) was used to select the most significant variables under the binary logistic regression analysis, with variable entry testing based on the significance of the score testing (at 0.05 level of significance), and the variable removal testing based on the probability of a likelihood ratio statistic.

4.3.5.1 Baseline model

Table 4.8 represents the results of baseline model. This model has only constant term. Logistic regression compares baseline model with a model that include all the significant factors to determine whether the latter model is more fitted.

Table 0.9: Baseline Model

Baseline Model	B	S.E.	Wald	df	Sig.	Exp(B)
Constant	-3.845	.047	6569.996	1	.000	.021

Initial -2 Log Likelihood (-2LL): 4408.544

Table 4.8, shows that coefficient of constant is -3.845 and standard error of the constant is 0.047. Wald Chi-Square statistic is tested the null hypothesis that the constant equal to 0. Above table shows that, constant is statistically significant to the model. Initial log likelihood value of the baseline model is 4408.544. The -2LL statistic is indicated how much information are unexplained from the model with large value of -2LL indicate lack of fitted model. This value is used to select an optimal model.

4.3.5.2 Developed model

Forward selection method is used to develop the binary logistic model. Variable which has minimum -2 log likelihood ratio is added first. The analysis was performed on p value equals to 0.05 significance level to formulate the model.

There are 13 variables were selected to test the association with poverty status of the household. First, all 13 variables are considered for develop the logistic regression model (result is in Appendix - D) and model iteration occurred up to nine steps. Residential Sector, Ethnicity of the household head, Education level of the household head, Marital status of the household head, Any of the household member receive income as an employee, Telephone facilities in the household area, Household size, Age of the household head and THHI have a significant effect on the poverty status of the household. But Hosmer and Lemeshow test (P value=0.001) indicates the model is lack of fit. Therefore, newest model was developed. It is identified that the issues of the 1st model is due to income of the household while highly significant to

the model. To develop best model for the poverty status, there is no option without removing the THHI from the model (Converting as a categorical variable also having same problem). Table 4.9 illustrate the finally developed model using forward selection method not considering income as a factor and model iteration occurred up to eight steps.

Table 0.10: Details of developed model

Variable	Factor	Coefficient (B)	S.E.	Wald	Df	P-value	Exp(B)
Residential Sector	Ru*			31.192	2	0.000	
	Ur*	-1.056	0.209	25.535	1	0.000	0.348
	Es*	-0.541	0.186	8.426	1	0.004	0.582
Ethnicity	Sin*	-0.836	0.109	59.051	1	0.000	0.433
Education level of the household head	PE*			116.895	3	0.000	
	OL*	-0.966	0.105	83.984	1	0.000	0.381
	AL*	-3.213	0.508	39.933	1	0.000	0.040
	HEdu*	-2.947	1.006	8.582	1	0.003	0.053
Telephone facilities in the household area	Tel_Y*	-0.571	0.111	26.347	1	0.000	0.565
Pipe borne line (main line) nearby household area	Wat_Y*	-0.259	0.108	5.704	1	0.017	0.772
Any Household member engage to agricultural activity	Agri_Y*	-0.415	0.115	13.090	1	0.000	0.660
Household size	HHS*	0.413	0.026	247.666	1	0.000	1.511
Age of the household head	Age*	-0.018	0.004	20.307	1	0.000	0.983
Constant*		-2.727	0.249	120.121	1	0.000	0.065

*Significant at 0.05 level of significance

Table 4.9 explains the variables in the developed model used to predict the poverty status of the household. When exploring results of this table, Residential Sector, Ethnicity of the household head, Education level of the household head, Telephone facilities in the household area, Pipe borne line (main line) nearby household area, Any Household member engage to agricultural activity, Household size and Age of the household head have a significant effect on the poverty status of the household. By observing the B (coefficients), it reveals that variables such as Residential Sector, Ethnicity of the household head, Education level of the household head, Telephone facilities in the household area, Pipe borne line (main line) nearby household area, Any Household member engage to agricultural activity and Age of the household head have a decreasing effect on the probability of a poverty status of the household. Only Household size has an increasing effect on the probability of a poverty status.

Gender of the household head, marital status of the household head, Any of the household member receive income as an employee and Household head suffer from chronic illness/disability are not statistically significant. That is, above mentioned variables are not much affect to the poverty status of the household.

Wald chi-square tests whether each of the predictors included make a significant contribution to the model while controlling other predictors.

Exp(B) is an OR. OR of urban sector indicates that urban sector households are about 1/3 as likely as a rural sector household to be poor household. That is urban sector households are 65.2% less like to be poor households than rural sector. Similarly, estate sector households are 41.8% less like to be poor households than rural sector. When we consider odd ratio of ethnicity of the household, it can be concluded that the Sinhala households are 56.7% less like to be poor households than other ethnicities.

ORs of the education levels indicate that households with education of the household head is Up to O/L, Up to A/L and Up to Higher education are 61.9%, 96.0% and 94.7% less like to be poor households than households with education of the household head is only up to Primary education respectively. These results indicate

that when the education level of the household head is high, chance to be a poor household is very low.

Households which have telephone facilities in their area are 43.5% less like to be poor households than the households which haven't telephone facilities in their area. Similarly, odds ratio of the Wat_Y indicates that households which has pipe borne line (main line) nearby their area are 22.3% less like to be poor households than the households which haven't pipe borne line (main line) nearby their area. These two results conclude that not having telephone facilities in the area is more associated to poor households than not having water facilities in the area.

OR of the Agri_Y indicates that households which any of the household members engaged to agricultural activity are 34.0% less like to be poor households than the households which no one of the household member engaged to agricultural activity. This result reveals that if household can earn money from any of the agricultural activity, it is more chance to be a non-poor household.

There are two continuous variables are in the model namely household size and age of the household head. Table 4.11 is indicated the Exp(B) associated with household size is 1.511. This indicates that for every one unit increase in household size (one additional member in the household), the odds of be a poor household increases by 51.1% which implies the higher the household size lead to higher the poverty risk. Odds ratio associated with age of the household is 0.983. This indicates that for every one unit increase in age of the household (one additional year of living), the odds of be a poor household decrease by 1.7% which implies the older household head leads to less the poverty risk in small proportion (0.017).

4.3.5.3 Importance of variables in the model

Table 4.10 presents the information how the model is affected if an explanatory variable is added to the model and in other word, which variable is important for the model.

Table 0.11: Description of the forward steps

Step	Model			Improvement			Added Variable
	-2LL	df	P value	Change in -2 LL	df	P value	
0	4408.544	1	0.000	-	-	-	-
1	4133.298	1	0.000	275.246	1	0.000	Household size
2	3914.993	4	0.000	218.304	3	0.000	Education level
3	3841.089	5	0.000	73.904	1	0.000	Ethnicity
4	3786.784	6	0.000	54.305	1	0.000	Telephone facilities in the household area
5	3744.561	8	0.000	42.223	2	0.000	Residential Sector
6	3721..816	9	0.000	22.745	1	0.000	Age of the household
7	3709.595	10	0.000	12.221	1	0.000	Any of the household member engaged to agricultural activity
8	3703.878	11	0.000	5.717	1	0.017	Pipe borne line (main line) nearby the household area

According to the results in table 4.10, adding the variable household size to the model makes the biggest change of the -2LL value of the model. Therefore, household size is the most important variable in this model. Similarly, adding the education level to the model also makes the biggest change of the -2LL value of the model. It is also identified that Pipe borne line (main line) nearby the household area is less important to the model than other variables in the model.

4.3.5.4 Measures of Goodness of Fit

Once a logistic model is fitted to the data it is essential to check that the assumed model is actually a valid model. Various measures are used to test the goodness of fit of the model.

a) Test of Model coefficients

The Omnibus test of model coefficient is checked the full model has an improvement over the baseline model. Chi-square statistic is used to test the significant difference between the Log-likelihoods (specifically the -2LLs) of the baseline model and the

full model. If the full model has significantly small -2LL value compared with baseline model, the full model is explaining more of the variance in the outcome and there is an improvement.

Table 0.12: Omnibus Tests of Model Coefficients

	Chi-square	df	p value
Step 8	5.717	1	0.017
Block	704.666	11	0.000
Model	704.666	11	0.000

Table 4.11 indicates that the model chi-square is highly significant (chi-square=704.666, p=0.000 with df =11). Thus, the developed model is significantly better than the baseline model. That means, the accuracy of the model improved when adding independent variables.

b) Model summary

Developed model is checked whether it is an improvement over the baseline model. Results are shown in the Table 4.12.

Table 0.13: Model summary

Step	-2 Log Likelihood	Cox and Snell R^2	Nagelkerke R^2
8	3703.878	.541	.573

According to the results of the Table 4.12, the developed model has a significantly small log likelihood value (3703.878) compared with the baseline model. It is revealed that the developed model is explaining more of the variance in the outcome and it is an improvement over the baseline model. Thus, it can be concluded that the developed model is better to predict the poverty of the household than the baseline model where no predictor variables were added. Also, Cox & Snell R^2 and Nagelkerke R^2 values are used to calculate the explained variation. According to these both values, the explained variation of the status of household poverty from the developed model is 54.1% and 57.3% respectively.

c) Predictive accuracy of the development model

Classification table is present the cross-classifying observed value of the dependent variable with a values of dichotomous variable derived from the logistic probabilities estimated by the developed model. To obtain the derived dichotomous variable first should define a cut point and compare each probability to the cut point. If the estimated probability is exceeded cut point then the derived variable equals to 1; otherwise it is equal to 0. Cut point can be obtained by plotting sensitivity (probability of detecting true positives) and specificity (probability of detecting true negatives) vs. all possible cut points.

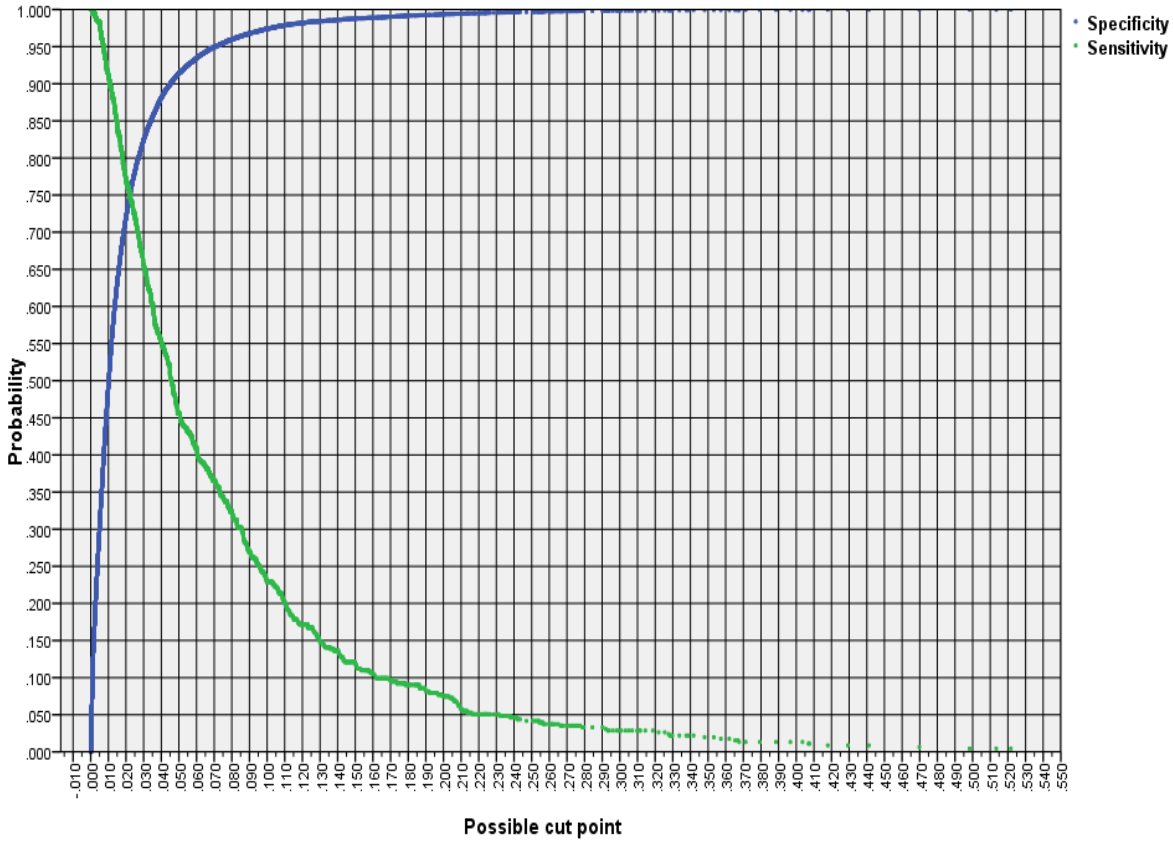

Figure 0.7: Plot of sensitivity and specificity Vs all possible cut point

According to the Figure 4.7, it can be identified that approximately sensitivity and specificity curves cross at 0.021. That is the optimal cut point. Table 4.13 shows that the classification of logistic model result by taking cut point equal to 0.021.

Table 0.14: Classification table

Observed	Predicted		
	Non-Poverty	Poverty	Percentage Correct
Non-Poverty	10276	3730	73.4
Poverty	72	228	76.0
Overall Percentage			73.4

Table 4.13 indicates that 76.0% were correctly classified for poverty household and 73.4% for non-poverty households. Overall 73.4% were correctly classified. It can be seen that the developed model is accurately classified the outcome for 73.4% of the cases compared to 62.3% in the null model.

d) Receiver Operating Characteristic Curve (ROC Curve)

Receiver operating characteristic (ROC) curve is used to evaluate the fit of a logistic regression model. This ROC curve plots the "sensitivity" and "1-specificity" for all possible cut points. ROC curve is shown in the Figure 4.8 and the area under the curve is presented in Table 4.14.

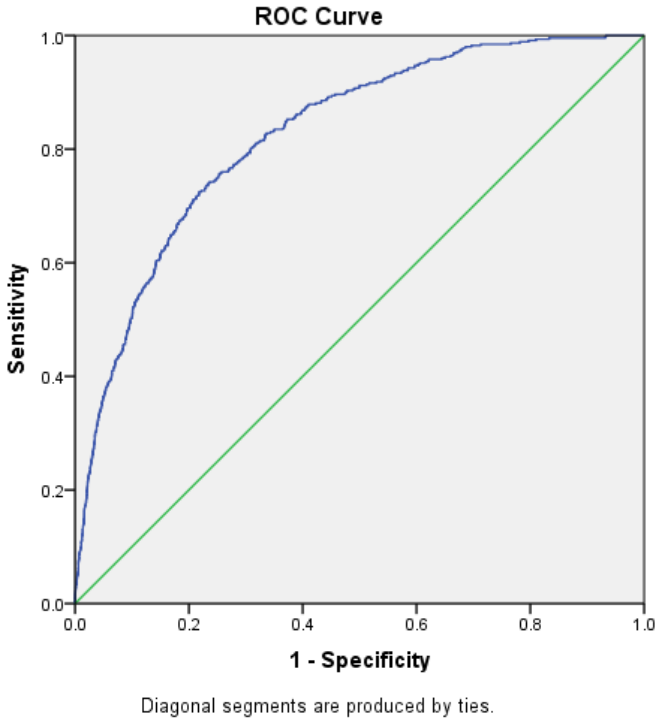

Figure 0.8: ROC Curve

Table 0.15: Details of the area under the ROC Curve

Area	Standard error	Significant value	95% Confidence Interval	
			Lower Bound	Upper Bound
0.827	0.009	0.000	0.809	0.845

According to the Table 4.14 and Figure 4.7, the area under the curve is 0.827 with 95% confidence interval (0.809, 0.845). The area under the curve is measured the accuracy of classification by the model. Since the P value = 0.000<0.05, it can be concluded that area under the curve is significantly differ from 0.5. That means, the logistic regression model is classified the group significantly better than by chance.

e) Hosmer and Lemeshow test

Hosmer and Lemeshow test are used to indicate which extent to the estimated model provides a better fit to the data than the null model. According to the table 4.15, Hosmer and Lemeshow test is concluded the model is a good fit to the data as P value = 0.461(>0.05).

Table 0.16: Hosmer and Lemeshow test

Chi-square	DF	Significant value
7.722	8	0.461

4.3.5.5 Model diagnostics

In case of more than two dependent variables in the logistic regression model, the standardized residual plots can be highlighted little regarding influential outliers (S.K.Sarkar, Midi, & Rana, 2011).

Different types of diagnostic plot have been identified to detect outliers and influence cases. Since it is in vain to consider all possible diagnostic plots, this study consider only 2 diagnostic plots more easily obtained and those are meaningful in logistic regression. Those are plotting Leverage value (Hi) and Change in parameter estimates ($\Delta\beta$) vs estimated logistic probability.

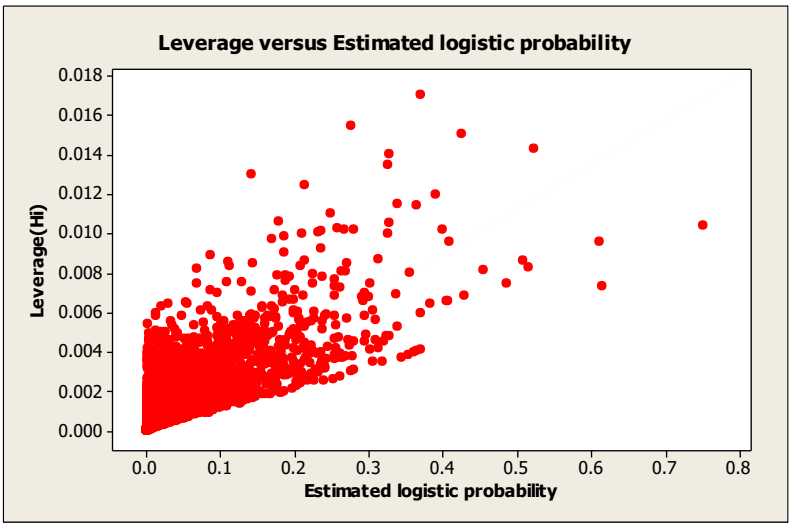

Figure 0.9: Plot of leverage Vs estimated logistic probability

The leverage value varies from 0 to 1 and high leverage values indicate the influential outliers. According to Figure 4.9, all leverage values are less than 1 even less than 0.02. Therefore, it can be identified that the outlying cases are not so influential.

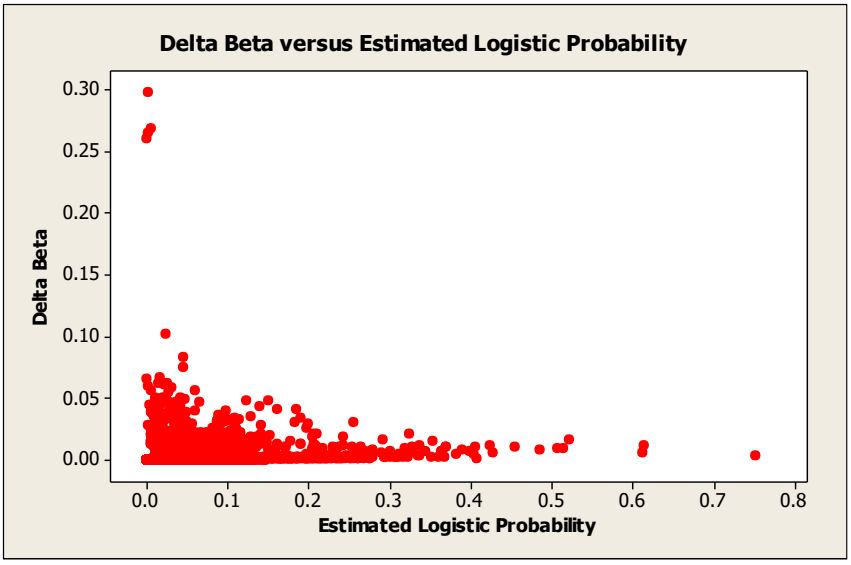

Figure 0.10: Plot of delta beta Vs estimated logistic probability

The effect on the set of parameter estimates in logistic regression when any specific observation is excluded can be observed using Plot of Delta Beta Vs Estimated Logistic Probability. According to Figure 4.10, all delta beta values are less than 1 even less than 0.3 indicates that the values are not large with respect to 1. This plot is also concluded that outlying cases are not so influential.

4.3.5.6 Model validation

Developed logistic regression model is used to validate the model. One third of data is used for validation and the results are shown in the Table 4.16.

Table 0.17: Classification table for model validation

Observed	Predicted		
	Non-Poverty	Poverty	Percentage Correct
Non-Poverty	5366	1848	74.4
Poverty	38	116	75.3
Overall Percentage			74.4

The model correctly predicted 74.4% of the validation data. That means the developed model accurately predicts the poverty of the household.

CHAPTER 5
CONCLUSIONS AND RECOMENDATIONS

5.1 Overviews

This chapter will explain the conclusions from the study and some recommendations are stated.

5.2 Conclusions

The main objective of the study was to identify the significant factors affecting for poverty status of the household in Sri Lanka. Based on this main objective, data was obtained from HIES-2016 conducted by DCS in Sri Lanka.

Analysis is categorized mainly as preliminary and fundamental analyses. Under preliminary analysis, frequency distributions and graphical analysis of the factors with poverty status of the household was performed. Under fundamental analyses, one sample proportion test is used to reduce some factor levels due to small proportions. Then Spearmen's rank correlation test and Chi-square test are performed to test the significant association between indicated factors and the poverty of the household. Finally, binary logistic regression was carried out due to dichotomous nature of dependent variable (i.e. 2 outcomes namely Poverty and Non-Poverty). The conclusions achieve from this study are summarized as below.

According to the descriptive analysis, it revealed that, 2.1% households are poor. Among them, most of the poor households are in rural residential sector. When we consider the poverty in province wise, Northern and Eastern province have more poor households than others. In Northern province, Kilinochchi have more poor households than the other district in the province. Also, Batticaloa have more poor households than all other districts. Cross tabulations of residential sector Vs status of poverty and education level of the household Vs status of poverty are revealed that residential sector and education level of the households can be affected to the status of poverty.

Under advanced analysis, according to the spearman's rank correlation test and Chi-square test, it was concluded that Residential Sector, Gender, Ethnicity, Marital Status, Education level, Any of the member in the household is engage to agricultural activity, Any of the household member Receive income as an employee, Household head suffer from chronic illness/Disability, Telephone facilities in household area, Household size, Age of the household head, Income of the household and Pipe borne line (main line) nearby household area are significantly associated with the poverty status of the household. Only three factors namely Main activity of the household head, Child death in the household and Electricity supply (main line) nearby household area are not significantly correlated with the poverty status. Therefore, these three non significant factors were not used to perform binary logistic regression analysis.

Based on the binary logistic regression analysis results, Residential Sector, Ethnicity of the household head, Education level of the household head, Telephone facilities in the household area, Pipe borne line (main line) nearby household area, Any Household member engage to agricultural activity and Age of the household head have a decreasing effect on the probability of a poverty status of the household and only Household size has an increasing effect on the probability of a poverty status of the household while gender of the household head, marital status of the household head, any of the household member receive income as an employee and household head suffer from chronic illness/disability are not statistically significant. This result is suggested that 3 of the household head characters namely Ethnicity, Age and education level are significant for poverty of the household. It also concludes that demographic characteristic like marital status and gender are not significantly affect to the poverty status of the household.

By considering OR following conclusions can be made.

- Rural residential sector households have more chance to be a poor than the other residential sector and second is estate sector.
- If Ethnicity of the household head is Sinhala, that household is less like to be poor households than other ethnicities.

- The education level of the household head is high, chance to be a poor household is very low.
- Not having telephone facilities in the area is more associated to poor households than not having water facilities in the area.
- Households which any of the household members engaged to agricultural activity are less like to be poor households than the households which no one of the household member engaged to agricultural activity. That is household can earn money from any of the agricultural activity, it is more chance to be a non-poor household.
- Higher the household size leads to higher the poverty risk.
- Older household head leads to less the poverty risk.

5.3 Recommendation

"End poverty in all its forms everywhere by 2030" is the first goal in SDG. To achieve this SDG, Sri Lankan government has to pay attention on the education of the people, utility facilities of the general public. Also, following recommendations are presented based on this study.

- It is chance to be a non-poor by increasing education level of the household. Therefore, government should make a plan to give education for all around the country.
- As like urban sector, government should pay more attention in Rural and Estate sector by giving utility facilities.
- Government should have plan a programme for helping households which has large household size.
- Northern and Eastern provinces have large number of households than the other province. Therefore, it is necessary to pay more attention on these two provinces to reduce the poverty in Sri Lanka.
- This study only considered main effect factors influence to the poverty of the household, it is better to perform research by considering interaction effects also.

REFERENCES

Achia, T. N., Wangombe, A., & Khadioli, N. (2010). A Logistic Regression Model to Identify Key Determinants of Poverty Using Demographic and Health Survey Data. European Journal of Social Sciences-Volume 13, 38-45.

Augustyn, A. (2019, 05 10). Poverty. Retrieved from Encyclopaedia Britannica : https://www.britannica.com/topic/poverty

Dagum, C., & Costa, M. (2004). Analysis and Measurement of Poverty: Univariate and Multivariate Approaches and Their Policy Implications, A Case Study Italy. Physica-Verlag Heidelberg.

Deepawansa, M., Sooriyarachchi, M., & Wickremasinghe, W. (n.d.). Determinants of Poverty in Sri Lanka. (pp. 340-344). Colombo: Department of Statistics, University of Colombo.

Denning, S. (2017, 08 15). Poverty Update: The Four Main Causes of Poverty in Sri Lanka. Retrieved from BORGEN Magazine: https://www.borgenmagazine.com/causes-of-poverty-in-sri-lanka/

Dhananjaya, S., & Alibuhtto, M. (2016). Factors Influencing Road Accidents in Sri Lanka: A Logistic Regression Approach. 157-173.

Duarte, R. (2016). Socioeconomic Determinants of Poverty in Paraguay. St. Louis: Washington University.

Dudek, H., & Lisicka, I. (2013). Determinants Of Poverty – Binary Logit Model with Interaction Terms Approach. Econometrics, 65-77.

Geda, A., Jong, N., Mwabu, G., & Kimenyi, M. (2001). Determinants of Poverty in Kenya: A Household Level Analysis. The Netherlands: ORPAS - Institute of Social Studies.

Gunewardena, D. (2005). Improving poverty measurement in Sri Lanka. Peradeniya: University of Peradeniya.

Harrell, F. E., Lee, K., & Mark, D. (1996). Tutorial in Biostatistics. Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. Statistics in Medicine, 361-387.

Hashmi, A. A., Hashmi, M. H., & Sial, A. H. (2008). Trends and Determinants of Rural Poverty:A Logistic Regression Analysis of Selected Districts of Punjab. The Pakistan Development Review, 909–923.

Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression. Hoboken: John Wiley & Sons.

Hosmer, D., Hosmer, T., Lemeshow, S., & Cessie, L. (1997). A comparison of goodness of fit tests for the logistic regression model. Statistics in medicine, 965-980.

Joshi, N. P., Keshav Lall, M., & Luni, P. (2012). Determinants of Income and Consumption Poverty in Far-Western Rural Hills of Nepal - A Binary Logistic Regression Analysis. Journal of Contemporary India Studies: Space and Society, Hiroshima University, 51-61.

Kumara, P. H., & Gunewardena, D. (2017). Disability and poverty in Sri Lanka: a household level analysis. Sri Lanka Journal of Social Sciences, 53-69.

Kumara, P. T. (2014). An Analysis of Poverty in Sri Lanka: the Causes of Poverty Reduction between 2007 and 2010.

Midi, H., Sarkar, S., & Rana, S. (2013). Collinearity diagnostics of binary logistic regression model. Journal of Interdisciplinary Mathematics, 253-267.

Mok, T., Gan, C., & Sanyal, A. (2007). The Determinants of Urban Household Poverty in Malaysia. Journal of Social Sciences, 190-196.

Osowole, O., Uba, E., & Ugbechie, R. (2012). On the Identification of Core Determinants of Poverty: A Logistic Regression Approach. Mathematical Theory and Modeling, 45-52.

(2017). Poverty Indicators: Household Income and Expenditure Survey - 2016, Department of Census and Statistics. Department of Census and Statistics.

Ramos, J. F., García-Guerra, A. K., Rodriguez, J. G., & Ramírez, G. M. (2015, April 6). Munich Personal RePEc Archive. Retrieved from https://mpra.ub.uni-muenchen.de/77743/

Ranathunga, S. P. (2011, 10 18). The determinants of household poverty in Sri Lanka: 2006/2007. MPRA Paper No. 34174.

S.K.Sarkar, Midi, H., & Rana, S. (2011). Detection of outliers and influential observations in binary logistic regression: An emperical study. Jurnal of Applied Sciences, 26-35.

Semasinghe, W. M. (2010). Dimensions of Rural Poverty in Sri Lanka: An attempt to identify the crucial aspects of well-being of rural poor. Journal of Social Sciences – Sri Lanka, 228-251.

Silva, I. D. (2008). Micro-level determinants of poverty reduction in Sri Lanka: a multivariate approach. International Journal of Social Economics, 140-158.

UNDP. (2019, 05 10). Sustainable Development Goals. Retrieved from United Nations Development Programme: http://www.lk.undp.org/content/srilanka/en/home/sustainable-development-goals/goal-1-no-poverty.html

Wang, Y., & Wang, B. (2016). Multidimensional poverty measure and analysis: a case study from Hechi City, China. Wang and Wang SpringerPlus.

Xhafaj, E., & Nurja, I. (2014). Determination of The Key Factors that Influence Poverty Through Econometric Models. European Scientific Journal, 65-72.

APPENDIX – A

Cross tabulation on poverty status by Province

			Poverty status		Total
			Non-Poverty	Poverty	
Province	Western	Count	4927	18	4945
		% within Province	99.6%	0.4%	100.0%
		% within Poverty status	23.1%	4.0%	22.7%
		% of Total	22.6%	0.1%	22.7%
	Central	Count	2703	59	2762
		% within Province	97.9%	2.1%	100.0%
		% within Poverty status	12.7%	13.0%	12.7%
		% of Total	12.4%	0.3%	12.7%
	Southern	Count	3145	48	3193
		% within Province	98.5%	1.5%	100.0%
		% within Poverty status	14.8%	10.5%	14.7%
		% of Total	14.5%	0.2%	14.7%
	Northern	Count	1893	111	2004
		% within Province	94.5%	5.5%	100.0%
		% within Poverty status	8.9%	24.4%	9.2%
		% of Total	8.7%	0.5%	9.2%
	Eastern	Count	1868	88	1956
		% within Province	95.5%	4.5%	100.0%
		% within Poverty status	8.8%	19.3%	9.0%
		% of Total	8.6%	0.4%	9.0%
	North Western	Count	2248	26	2274
		% within Province	98.9%	1.1%	100.0%
		% within Poverty status	10.6%	5.7%	10.5%
		% of Total	10.3%	0.1%	10.5%

	North Central	Count	1347	17	1364
		% within Province	98.8%	1.2%	100.0%
		% within Poverty status	6.3%	3.7%	6.3%
		% of Total	6.2%	0.1%	6.3%
	Uva	Count	1278	37	1315
		% within Province	97.2%	2.8%	100.0%
		% within Poverty status	6.0%	8.1%	6.0%
		% of Total	5.9%	0.2%	6.0%
	Sabaragamuwa	Count	1892	51	1943
		% within Province	97.4%	2.6%	100.0%
		% within Poverty status	8.9%	11.2%	8.9%
		% of Total	8.7%	0.2%	8.9%
Total		Count	21301	455	21756
		% within Province	97.9%	2.1%	100.0%
		% within Poverty status	100.0%	100.0%	100.0%
		% of Total	97.9%	2.1%	100.0%

APPENDIX – B

Cross tabulation on poverty status by District

			Poverty status		Total
			Non-Poverty	Poverty	
District	Colombo	Count	1981	5	1986
		% within District	99.7%	0.3%	100.0%
		% within Poverty status	9.3%	1.1%	9.1%
		% of Total	9.1%	0.0%	9.1%
	Gampaha	Count	1809	6	1815
		% within District	99.7%	0.3%	100.0%
		% within Poverty status	8.5%	1.3%	8.3%
		% of Total	8.3%	0.0%	8.3%
	Kalutara	Count	1137	7	1144
		% within District	99.4%	0.6%	100.0%
		% within Poverty status	5.3%	1.5%	5.3%
		% of Total	5.2%	0.0%	5.3%
	Kandy	Count	1279	36	1315
		% within District	97.3%	2.7%	100.0%
		% within Poverty status	6.0%	7.9%	6.0%
		% of Total	5.9%	0.2%	6.0%
	Matale	Count	623	11	634
		% within District	98.3%	1.7%	100.0%
		% within Poverty status	2.9%	2.4%	2.9%
		% of Total	2.9%	0.1%	2.9%
	Nuwaraeliya	Count	801	12	813
		% within District	98.5%	1.5%	100.0%
		% within Poverty status	3.8%	2.6%	3.7%
		% of Total	3.7%	0.1%	3.7%
	Galle	Count	1240	14	1254
		% within District	98.9%	1.1%	100.0%
		% within Poverty status	5.8%	3.1%	5.8%
		% of Total	5.7%	0.1%	5.8%
	Matara	Count	1111	30	1141
		% within District	97.4%	2.6%	100.0%
		% within Poverty status	5.2%	6.6%	5.2%
		% of Total	5.1%	0.1%	5.2%

	Hambantota	Count	794	4	798
		% within District	99.5%	0.5%	100.0%
		% within Poverty status	3.7%	0.9%	3.7%
		% of Total	3.6%	0.0%	3.7%
	Jaffna	Count	661	26	687
		% within District	96.2%	3.8%	100.0%
		% within Poverty status	3.1%	5.7%	3.2%
		% of Total	3.0%	0.1%	3.2%
	Mannar	Count	324	3	327
		% within District	99.1%	0.9%	100.0%
		% within Poverty status	1.5%	0.7%	1.5%
		% of Total	1.5%	0.0%	1.5%
	Vavuniya	Count	339	4	343
		% within District	98.8%	1.2%	100.0%
		% within Poverty status	1.6%	0.9%	1.6%
		% of Total	1.6%	0.0%	1.6%
	Mullaittivu	Count	267	32	299
		% within District	89.3%	10.7%	100.0%
		% within Poverty status	1.3%	7.0%	1.4%
		% of Total	1.2%	0.1%	1.4%
	Killinochchi	Count	302	46	348
		% within District	86.8%	13.2%	100.0%
		% within Poverty status	1.4%	10.1%	1.6%
		% of Total	1.4%	0.2%	1.6%
	Batticaloa	Count	655	48	703
		% within District	93.2%	6.8%	100.0%
		% within Poverty status	3.1%	10.5%	3.2%
		% of Total	3.0%	0.2%	3.2%
	Ampara	Count	754	12	766
		% within District	98.4%	1.6%	100.0%
		% within Poverty status	3.5%	2.6%	3.5%
		% of Total	3.5%	0.1%	3.5%
	Trincomalee	Count	459	28	487
		% within District	94.3%	5.7%	100.0%
		% within Poverty status	2.2%	6.2%	2.2%
		% of Total	2.1%	0.1%	2.2%

	Kurunagala	Count	1529	17	1546
		% within District	98.9%	1.1%	100.0%
		% within Poverty status	7.2%	3.7%	7.1%
		% of Total	7.0%	0.1%	7.1%
	Puttalam	Count	719	9	728
		% within District	98.8%	1.2%	100.0%
		% within Poverty status	3.4%	2.0%	3.3%
		% of Total	3.3%	0.0%	3.3%
	Anuradhapura	Count	783	10	793
		% within District	98.7%	1.3%	100.0%
		% within Poverty status	3.7%	2.2%	3.6%
		% of Total	3.6%	0.0%	3.6%
	Polannaruwa	Count	564	7	571
		% within District	98.8%	1.2%	100.0%
		% within Poverty status	2.6%	1.5%	2.6%
		% of Total	2.6%	0.0%	2.6%
	Baddulla	Count	739	23	762
		% within District	97.0%	3.0%	100.0%
		% within Poverty status	3.5%	5.1%	3.5%
		% of Total	3.4%	0.1%	3.5%
	Monaragala	Count	539	14	553
		% within District	97.5%	2.5%	100.0%
		% within Poverty status	2.5%	3.1%	2.5%
		% of Total	2.5%	0.1%	2.5%
	Rattnapura	Count	1015	27	1042
		% within District	97.4%	2.6%	100.0%
		% within Poverty status	4.8%	5.9%	4.8%
		% of Total	4.7%	0.1%	4.8%
	Kegalle	Count	877	24	901
		% within District	97.3%	2.7%	100.0%
		% within Poverty status	4.1%	5.3%	4.1%
		% of Total	4.0%	0.1%	4.1%
Total		Count	21301	455	21756
		% within District	97.9%	2.1%	100.0%
		% within Poverty status	100.0%	100.0%	100.0%
		% of Total	97.9%	2.1%	100.0%

APPENDIX – C

Normality test for continuous variables with Poverty Status

	Poverty status	Kolmogorov-Smirnov[a]		
		Statistic	df	Sig.
Total with windfall	Non-Poverty	.388	21301	.000
	Poverty	.205	455	.000
Household size	Non-Poverty	.141	21301	.000
	Poverty	.143	455	.000
Age	Non-Poverty	.042	21301	.000
	Poverty	.069	455	.000

APPENDIX – D

Logistic regression model with 13 variables including Household Income

Block 0: Beginning Block

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-3.845	.047	6569.996	1	.000	.021

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
	Step	275.246	1	.000
Step 1	Block	275.246	1	.000
	Model	275.246	1	.000
	Step	317.446	1	.000
Step 2	Block	592.692	2	.000
	Model	592.692	2	.000
	Step	128.123	3	.000
Step 3	Block	720.815	5	.000
	Model	720.815	5	.000
	Step	50.939	1	.000
Step 4	Block	771.754	6	.000
	Model	771.754	6	.000
	Step	31.145	1	.000
Step 5	Block	802.899	7	.000
	Model	802.899	7	.000
	Step	33.289	1	.000
Step 6	Block	836.189	8	.000
	Model	836.189	8	.000
	Step	35.906	2	.000
Step 7	Block	872.095	10	.000
	Model	872.095	10	.000
	Step	19.661	1	.000
Step 8	Block	891.756	11	.000
	Model	891.756	11	.000
	Step	4.499	1	.034
Step 9	Block	896.255	12	.000
	Model	896.255	12	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4133.298[a]	.013	.069
2	3815.852[b]	.027	.147
3	3687.728[c]	.033	.178
4	3636.790[c]	.035	.190
5	3605.644[c]	.036	.198
6	3572.355[c]	.038	.206
7	3536.449[c]	.039	.214
8	3516.788[c]	.040	.219
9	3512.289[c]	.041	.220

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

b. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

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

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5.552	4	.235
2	23.750	8	.003
3	21.180	8	.007
4	17.926	8	.022
5	23.962	8	.002
6	29.841	8	.000
7	24.838	8	.002
8	22.255	8	.004
9	25.322	8	.001

Classification Table[a]

	Observed		Predicted		
			Poverty status		Percentage Correct
			Non-Poverty	Poverty	
Step 1	Poverty status	Non-Poverty	14729	6491	69.4
		Poverty	173	281	61.9
	Overall Percentage				69.3
Step 2	Poverty status	Non-Poverty	16729	4491	78.8
		Poverty	144	310	68.3
	Overall Percentage				78.6
Step 3	Poverty status	Non-Poverty	16611	4609	78.3
		Poverty	117	337	74.2
	Overall Percentage				78.2
Step 4	Poverty status	Non-Poverty	16716	4504	78.8
		Poverty	116	338	74.4
	Overall Percentage				78.7
Step 5	Poverty status	Non-Poverty	16836	4384	79.3
		Poverty	113	341	75.1
	Overall Percentage				79.3
Step 6	Poverty status	Non-Poverty	16853	4367	79.4
		Poverty	114	340	74.9
	Overall Percentage				79.3
Step 7	Poverty status	Non-Poverty	16914	4306	79.7
		Poverty	102	352	77.5
	Overall Percentage				79.7
Step 8	Poverty status	Non-Poverty	16903	4317	79.7
		Poverty	100	354	78.0
	Overall Percentage				79.6
Step 9	Poverty status	Non-Poverty	16885	4335	79.6
		Poverty	107	347	76.4
	Overall Percentage				79.5

a. The cut value is .025

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Householdsize	.434	.025	305.481	1	.000	1.543	1.470	1.620
	Constant	-5.754	.134	1852.383	1	.000	.003		
Step 2[b]	Householdsize	.613	.027	506.331	1	.000	1.846	1.750	1.948
	THHI	-7.416	.403	339.228	1	.000	.001	.000	.001
	Constant	7.238	.684	111.973	1	.000	1391.395		
Step 3[c]	Recode_education			89.998	3	.000			
	Recode_education(1)	-.777	.100	60.888	1	.000	.460	.378	.559
	Recode_education(2)	-2.943	.506	33.784	1	.000	.053	.020	.142
	Recode_education(3)	-2.233	1.006	4.922	1	.027	.107	.015	.771
	Householdsize	.578	.027	443.090	1	.000	1.783	1.690	1.882
	THHI	-6.703	.429	244.431	1	.000	.001	.001	.003
	Constant	6.638	.722	84.503	1	.000	763.805		
Step 4[d]	Recode_education			121.472	3	.000			
	Recode_education(1)	-.991	.104	90.026	1	.000	.371	.303	.456
	Recode_education(2)	-3.219	.508	40.155	1	.000	.040	.015	.108
	Recode_education(3)	-2.435	1.007	5.852	1	.016	.088	.012	.630
	Householdsize	.582	.028	426.837	1	.000	1.789	1.693	1.890
	Age	-.026	.004	49.285	1	.000	.974	.967	.981
	THHI	-6.805	.437	242.037	1	.000	.001	.000	.003
	Constant	8.285	.774	114.435	1	.000	3964.143		
Step 5[e]	Recode_ethnicity(1)	-.578	.103	31.225	1	.000	.561	.458	.687
	Recode_education			101.480	3	.000			
	Recode_education(1)	-.899	.105	72.619	1	.000	.407	.331	.501
	Recode_education(2)	-3.067	.509	36.341	1	.000	.047	.017	.126
	Recode_education(3)	-2.365	1.007	5.518	1	.019	.094	.013	.676
	Householdsize	.547	.029	366.677	1	.000	1.727	1.633	1.827
	Age	-.022	.004	33.113	1	.000	.978	.971	.986
	THHI	-6.503	.442	216.156	1	.000	.001	.001	.004
	Constant	7.944	.781	103.464	1	.000	2817.554		
Step 6[f]	Recode_ethnicity(1)	-.624	.104	35.848	1	.000	.536	.437	.657
	Recode_education			85.316	3	.000			
	Recode_education(1)	-.823	.106	59.696	1	.000	.439	.357	.541
	Recode_education(2)	-2.907	.510	32.552	1	.000	.055	.020	.148
	Recode_education(3)	-2.196	1.008	4.747	1	.029	.111	.015	.802
	Is_Tel_Lines_Near(1)	-.616	.104	34.843	1	.000	.540	.440	.663
	Householdsize	.552	.029	370.219	1	.000	1.736	1.641	1.836
	Age	-.020	.004	26.342	1	.000	.981	.973	.988

	THHI	-6.351	.446	202.830	1	.000	.002	.001	.004
	Constant	7.935	.786	101.866	1	.000	2793.616		
	New_Sector			31.036	2	.000			
	New_Sector(1)	-.962	.206	21.775	1	.000	.382	.255	.572
	New_Sector(2)	-.667	.189	12.459	1	.000	.513	.354	.743
	Recode_ethnicity(1)	-.802	.108	54.792	1	.000	.449	.363	.555
	Recode_education			85.154	3	.000			
	Recode_education(1)	-.831	.107	60.068	1	.000	.436	.353	.538
Step 7[g]	Recode_education(2)	-2.898	.510	32.280	1	.000	.055	.020	.150
	Recode_education(3)	-2.155	1.009	4.568	1	.033	.116	.016	.836
	Is_Tel_Lines_Near(1)	-.528	.106	24.869	1	.000	.590	.479	.726
	Householdsize	.563	.029	375.596	1	.000	1.756	1.659	1.859
	Age	-.019	.004	24.466	1	.000	.981	.974	.989
	THHI	-6.268	.446	197.319	1	.000	.002	.001	.005
	Constant	7.882	.790	99.478	1	.000	2648.448		
	New_Sector			35.927	2	.000			
	New_Sector(1)	-.982	.207	22.579	1	.000	.375	.250	.562
	New_Sector(2)	-.793	.191	17.262	1	.000	.452	.311	.658
	Recode_ethnicity(1)	-.806	.108	55.203	1	.000	.447	.361	.553
	Recode_education			78.502	3	.000			
	Recode_education(1)	-.791	.107	54.297	1	.000	.454	.368	.560
	Recode_education(2)	-2.824	.510	30.655	1	.000	.059	.022	.161
Step 8[h]	Recode_education(3)	-2.167	1.009	4.616	1	.032	.114	.016	.827
	IncomeASemployee(1)	.532	.124	18.486	1	.000	1.703	1.336	2.171
	Is_Tel_Lines_Near(1)	-.534	.106	25.404	1	.000	.586	.476	.721
	Householdsize	.556	.029	360.433	1	.000	1.743	1.646	1.846
	Age	-.017	.004	19.778	1	.000	.983	.975	.990
	THHI	-6.984	.492	201.843	1	.000	.001	.000	.002
	Constant	8.773	.839	109.250	1	.000	6460.154		
	New_Sector			35.475	2	.000			
	New_Sector(1)	-.965	.207	21.804	1	.000	.381	.254	.571
	New_Sector(2)	-.800	.191	17.519	1	.000	.449	.309	.654
	Recode_ethnicity(1)	-.812	.108	56.021	1	.000	.444	.359	.549
	Recode_education			80.626	3	.000			
Step 9[i]	Recode_education(1)	-.805	.107	56.195	1	.000	.447	.362	.552
	Recode_education(2)	-2.843	.510	31.058	1	.000	.058	.021	.158
	Recode_education(3)	-2.196	1.009	4.738	1	.030	.111	.015	.804
	Recode_maritalStatus(1)	-.298	.143	4.306	1	.038	.743	.561	.984
	IncomeASemployee(1)	.536	.124	18.683	1	.000	1.708	1.340	2.178

	Is_Tel_Lines_Near(1)	-.530	.106	25.054	1	.000	.588	.478	.724
	Householdsize	.547	.030	339.267	1	.000	1.728	1.630	1.832
	Age	-.015	.004	14.258	1	.000	.985	.977	.993
	THHI	-7.081	.495	204.547	1	.000	.001	.000	.002
	Constant	8.944	.847	111.560	1	.000	7663.730		