CO-OPERATIVE UNIVERSITY, SAGAING
DEPARTMENT OF STATISTICS
MASTER OF APPLIED STATISTICS

FACTORS AFFECTING ON POVERTY OF HOUSEHOLDS IN KYAUKPADAUNG TOWNSHIP

SU PAN HNIN

JULY, 2021

CO-OPERATIVE UNIVERSITY, SAGAING

DEPARTMENT OF STATISTICS

MASTER OF APPLIED STATISTICS

FACTORS AFFECTING ON POVERTY OF HOUSEHOLDS IN KYAUKPADAUNG TOWNSHIP

SU PAN HNIN

2MAS-1

JULY, 2021

FACTORS AFFECTING ON POVERTY OF HOUSEHOLDS IN KYAUKPADAUNG TOWNSHIP

This thesis is submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of Master of Applied Statistics.

Supervised by:

Daw Yin Mon Thant
Associate Professor
Department of Statistics
Co-operative University, Sagaing

Submitted by:

Ma Su Pan Hnin
2MAS-1
Master of Applied Statistics
Co-operative University, Sagaing

ACCEPTANCE

This is to certify that this paper entitled **"Factors Affecting on Poverty of Households in Kyaukpadaung Township"** submitted as a partial fulfillment towards the degree of Master of Applied Statistics has been accepted by Board of Examiners.

BOARD OF EXAMINERS

(Chairman)
Prof. Dr. Moe Moe Yee
Rector
Co-operative University, Sagaing

(External Examiner)
Prof. Dr. Kyaing Kyaing Thet
Pro-Rector
Monywa University of Economics

(External Examiner)
Prof. Daw Khin Aye Myint
Visiting Professor
Department of Statistics
Co-operativeUniversity,Sagaing

(Examiner)
Prof. Daw Khin San Kyi
Head of Department of Statistics
Co-operative University, Sagaing

(Examiner & Supervisor)
Associate Prof. Daw Yin Mon Thant
Department of Statistics
Co-operative University, Sagaing

JULY, 2021

ABSTRACT

The study is concerned with the factors affecting on poverty of households in Kyaukpadaung Township. The objectives of the study are to study the socio and economic situations of households, to analyze the factors affecting on the expenditure of households and to analyze the effect of socio-demographic and economic factors on poverty of households in Kyaukpadaung Township. The first stage is randomly selected four wards from 12 wards of Kyaukpadaung Township and the second stage is randomly collected 300 households from four wards by using simple random sampling method. Descriptive analysis, multiple regression analysis, principle component analysis using wealth index and binary logistic regression are applied from primary data. Most of the household heads are middle educational level. Income of households per month are usually 90000-374999 Kyats. In Multiple regression analysis, monthly income, number of students and household heads age are statistically significant. In Binary logistic regression, the coefficient of gender, occupation, total income and loans are statistically significant. Total income is not practically significant, although statistically significant. Therefore, the government should focus and support job opportunities to increase the income level and uplift the socio-economic life of household in Kyaukpadaung Township.

ACKNOWLEDGMENTS

First of all, I would like to express my deep appreciation and gratitude to Rector, Dr. Moe Moe Yee of Co-operative University, Sagaing, for her supportive advices, constructive comments and suggestions.

I would like to express my indebtedness Pro-Rector, Dr. Kyaing Kyaing Thet, Monywa University of Economics for her valuable comments and suggestions.

My special thanks go to Professor (Retired) Head of Department of Statistics, Daw Khin Aye Myint, who gave her valuable comments, wise guidance.

I would like to express my indebtedness to Daw Khin San Kyi, Professor and Head of Department of Statistics, for her permission to write thesis in this field of the study.

I am very deep thankful to my supervisor Daw Yin Mon Thant, Associate Professor, Department of Statistics, for her giving countless hours of her valuable time supervision and well understand my difficulties throughout my work.

In addition, I am profoundly grateful to all teachers in Department of Statistics, for their help and support.

Finally, I would like to extend my gratitude to my parent for allowing me to study the degree of Master of Applied Statistics and easy life.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
CCA	Canonical Correlation Analysis
EFA	Exploratory Factor Analysis
GDP	Gross Domestic Product
HH	Household Head
KMO	Kaiser-Meyer-Olkin
MEPI	Multidimensional Energy Poverty Index
MLCS	Myanmar Living Conditions Survey
MMK	Myanmar Kyat
MPI	Multidimensional Poverty Index
OPHI	Oxford Poverty and Human Development Initiative
PCA	Principal Component Analysis
PCR	Principal Components Regression
PLS	Partial Least Squares
RC	Reinforced Concrete
SHINE	Lessons from the Sanitation, Hygiene, Infant Nutrition Efficacy
S/N	States and Regions
TV	Television
UN	United Nation
VCD	Versatile Compact Discs
VIF	Variance Inflation Factor

CHAPTER 1
INTRODUCTION

There is nobody reason for poverty, and the results of it are different in every situation. Poverty varies considerably relying upon the circumstance. The differences between rich and poor within the borders of a country can similarly be amazing. Admittance to great schools, medical care, electricity, safe water, and other critical services stays slippery for some and is frequently controlled by financial status, gender, nationality, and geography. For ready to move out of poverty, progress is frequently temporary. Economic shocks, food weakness, and environmental change compromise the benefits and may drive a back into poverty.

Poverty is a troublesome cycle to break and regularly passed from one generation to the next. Common outcomes of poverty incorporate alcohol, substance abuse less, access to education, poor housing and living conditions, and expanded degrees of disease. Uplifted poverty is probably going to cause expanded pressures in society, as inequality increases. Like this issues frequently lead to increasing crime percentages in communities because of the influence of poverty.

Poverty can be decreased by increasing the income of the poor, or by reducing the expense of goods and services of consuming. It should be noticed that a fixed monetary poverty line could rapidly lead to the assumption that only income increases of poor households can lessen poverty. However, poverty can be multi-facetted, and a poor household couldn't care less about the absolute amount of receiving income, however about the purchase power of this income.

1.1 Rationale of the Study

The worldwide extreme poverty rate fell 10.1 percent to 9.2 percent from 2015 to 2017. That is identical to 689 million people living on less than $1.90 a day. At higher poverty lines, 24.1 percent of the world lived on under $3.20 a day and 43.6 percent on under $5.50 per day in 2017.

In 2018, four out of five people below the international poverty line lived in country areas. In most regions and among some age groups, women represent a majority of the poor. About 70 percent of the world poor aged 15 and over have no schooling or only some basic education. Almost half of poor people in Sub-Saharan Africa live in

only five nations: Nigeria, the Democratic Republic of Congo, Tanzania, Ethiopia, and Madagascar.

In excess of 40 percent of the worldwide poor live in economies influenced by delicacy, struggle and viciousness, and that number is relied upon to ascend to 67 percent in the following decade. Such economies have just 10 percent of the world's population. Around 132 million of the worldwide poor live in zones with high flood hazard. Middle-income countries for instance India and Nigeria will be essentially influenced; middle-income countries might be home to 82 percent of the new poor.

New exploration appraises that environmental change will drive 68 million to 132 million into poverty by 2030. Environmental change is an especially intense danger for nations in Sub-Saharan Africa and South Asia, the districts where the vast majority of the worldwide poor are concentrated. In various nations, an enormous portion of the poor live in regions that are both influenced by conflict and confronting high presentation to floods, for instance, Nepal, Cameroon, Liberia, and the Central African Republic.

Global extreme poverty is relied upon to ascend in 2020 and up to now for the first time in over 20 years as the disruption of the COVID-19 pandemic compounds the forces of conflict and environmental change, which were already slowing poverty decrease progress.

Myanmar has an expected population of 51.4 million, containing of various ethnic groups speaking over 100 languages and dialects. It is ranked 150 out of 187 nations on the Human Development Index. Economic development has averaged 5 percent in recent years with a per capita income of USD$702. Poverty levels are at an estimated 26 percent of the population. Poverty is twice as high in rural areas where 70 percent of the population lives.

Myanmar's poverty is larger, however shallow, with the median income only 25 percent above the poverty line. Small improvements can thus bring an enormous number of people out of poverty, but already small shocks can also bring an even larger number of people into poverty.

Poverty has always been a matter of concern for policy makers and social researchers. It limits the people's access to basic necessities of life such as food, shelter and clothing and also creates a sense of deprivation among people. It denies the fundamental right to act and chose freely which would enjoy if not poor. Poor people

are generally exposed to the exploitation by state and society. The people often lack political power and have less say in the decision making process that had direct impact on lives. People are more vulnerable to economic shocks such as fluctuations in economic growth and unprecedented rise in inflation. Thus, in case of any economic crises and fatalities, people lose more and pay higher prices than the rich. On the other hand, in good times of economic prosperity, people generally gain less as compared to rich. Most official statistics on poverty computed in rich and poor countries are based solely on household income.

Socio-demographic and economic factors for the poverty level in the household is one of the important facts. Therefore it is the study of socio-economic factors and poverty level of household in Kyaukpadaung Township.

1.2 Objectives of the Study

The objectives of the study are;

(i) to study the socio and economic situations of household in Kyaukpadaung Township.

(ii) to analyze the factors affecting on the expenditure of households.

(iii) to analyze the effect of socio-demographic and economic factors on poverty of households.

1.3 Method of Study

The main sources of data are primary using structural questionnaires with two stage random sampling methods. The first stage is randomly selected four wards from 12 wards of Kyaukpadaung Township and the second stage is randomly collected 300 households from four wards by using simple random sampling method. Descriptive analysis, multiple regression analysis, principle component analysis using wealth index and binary logistic regression are used in this study.

1.4 Scope and Limitations of the Study

This study focuses on the socio-economic status in poverty level of Aung Mingalar ward, Thar Yar Aye ward, Pyi Tharyar ward and Thiri Mingalar ward in Kyaukpadaung Township, January 2021 survey. This study was limited to all the 300 households to four wards in total twelve wards of Kyaukpadaung Township. There are

several ways to study poverty but in this study only descriptive analysis, multiple regression analysis, wealth index using principle component analysis and binary logistic regression are used. In this study included limited scarce funds to cover all the costs and difficult to face to face interview due to Covid-19 pandemic.

1.5 Organization of the Study

The study is organized into five chapters. Chapter 1 is the introduction which presents five sub-titles such as rationale of the study, objectives of the study, methods of study, scope and limitation and organization of the study. Chapter 2 covers literature review which includes into theoretical review and empirical review. Chapter 3 presents methodology using in this study. Chapter 4 constitutes analysis of factors affecting on poverty level in Kyaukpadaung Township. Chapter 5 discusses the conclusion based on findings, recommendations and suggestions.

CHAPTER 2
LITERATURE REVIEW

In this chapter, theoretical review and empirical reviews of concerning the poverty are presented.

2.1 Theoretical Review

The definition of poverty, history of poverty and history of poverty line in Myanmar are including in the theoretical review.

2.1.1 Definition of Poverty

Basically, poverty refers to lacking enough assets to give the necessities of life such as food, clean water, safe house and clothing. In any case, in the day and age, that can be reached out to include access to medical care, training and even transportation. In government circles, poverty is regularly characterized as “absolute poverty " and "relative poverty”.

Each nation has own measure for poverty. The Bank keeps a measurement called the International Poverty Line and, starting at 2015, set the meaning of extreme poverty as the individuals who live on under US$1.90 every day. The people who living on between $1.90- $3.10 every day are classified the "moderate poor. This number depends on the financial estimation of an individual's consumption instead of income alone.

The Multidimensional Poverty Index (MPI) is upheld by the United Nations Development Program. It estimates poverty across three measurements including health, education and standard of living. It at that point further separates into 10 indicators are nutrition, child mortality, years of schooling, school attendance, cooking fuel, sanitation, drinking water, electricity, housing and assets.

2.1.2 Absolute Poverty and Relative Poverty

Absolute poverty refers to people whose salaries fall under a line set by a given nation. Below this line individuals can't meet the fundamental requirements for food, water and shelter. The people, who fall under the line given by nation, also have no access to social administrations such as medical care, education and utilities.

Relative poverty refers to individuals whose total incomes are not exactly a specific rate, regularly half, of the nation's median income. Since the median income

can shift because of economic growth, the line for relative poverty can change. At the point when poverty is characterized to include access to services and security basic to well-being and not just income and consumption, the worldwide poverty rate increments by half.

2.1.3 History of Poverty

In the previous 200 years, the world has gained enormous ground in closure worldwide poverty. In 1820, by far most of the world lived in extreme poverty 200 years ago. Just a little tip top portion enjoyed higher standards of living. From that point forward, economic growth has changed our reality, lifting more individuals out of poverty even while population numbers have multiplied sevenfold. In 1945, following World War II, agents of 50 nations signed the United Nation Charter, which recognized that keeping up peace is associated with improved social development and social justice.

In 1964, President Lyndon Johnson announced “war on poverty “in the United States. In 1970, the quantity of individuals living in extreme poverty peaked at 2.2 billion. In 1981, the World Bank started collecting data on global poverty. Generally through household surveys found that 44 percent of the world lived in extreme poverty. In 1990, the World Bank characterized extreme poverty as people living on $1 or less a day. Around 1.89 billion people, or almost 36 percent of the total population, lived in extreme poverty. Almost a large portion of the population in agricultural nations lived on under $1.25 every day.

In 1992, the United Nation embraced Agenda 21, resolving to cooperate to battle global poverty utilizing country specific solutions. In 1995, the United Nations united the biggest get-together of world leaders until then, at the World Summit for Social Development, where leaders wrote the Copenhagen Declaration as a vow to kill poverty. In 1997, the United Nation General Assembly proclaimed the First U.N. Decade for Eradication of Poverty from 1997 to 2006, taking the responsibility from the Copenhagen Declaration and setting it into action. In 2000, all 191 United Nations member states signed the Millennium Development Goals, eight objectives to accomplish by 2015, including decreasing extreme poverty rates, at that point determined as individuals living on under $1 per day, by half.

In 2008, the World Bank restored the international poverty line as individuals living on $1.25 every day, utilizing 2005 prices for the cost of living. The leaders of 8 United Nation declared the Second U.N. Decade for Eradication of Poverty from 2008 to 2017, developing the success of the first decade and focusing on jobs and income generation as an approach to battle poverty. In 2010, the Millennium Development Goal of decreasing the 1990 extreme poverty rates significantly was accomplished five years sooner than anticipated.

In 2012, the U.N. General Assembly received another goal about the future of need, recognizing that, "Killing poverty is the best global challenge facing the present world." In 2015, the World Bank raised the international poverty line from $1.25 per day to $1.90, in view of 2011 prices for the cost of living. Additionally, United Nations member states received the Sustainable Development Goals, which include goals to end poverty and hunger in all structures. In 2020, global poverty is relied upon to ascend without precedent for a very long time. The COVID-19 pandemic threatens to push 88 million to 115 million individuals into extreme poverty in 2020, with the absolute ascending to upwards of 150 million by 2021.

The world has made enormous walks in overcoming global poverty. Since 1990, more than 1.2 billion people have emerged from extreme poverty. Presently, 9.2 percent of the world gets on $1.90 per day or less, compared to nearly 36 percent in 1990. At the point when families move out of poverty, children's health and wellbeing improve. Since 1990, the number of children under the age of 5 is died because of poverty, hunger and illness.

2.1.4 History of Poverty Line in Myanmar

A poverty line, additionally called a poverty threshold, is the line below which it is difficult, if not impossible, to manage the cost of fundamental requirements. The poverty line is determined in every nation by adding up the expense of meeting minimum needs, such as food and shelter. Household incomes that is too low to even consider affording least requirements, for example food and shelter, are below the poverty line.

The income necessary to manage meeting minimum needs normally sets the poverty line for a nation. Poverty lines can then be compared between nations. The international poverty line is the standard poverty line for estimating poverty universally.

Nonetheless, generally new estimates, for example, the Global 9 Multidimensional Poverty Index includes measurements of health, education, and living standards, all as indications of poverty.

The international poverty line, at present set at $1.90 every day, is the universal norm for estimating global poverty. This line helps measure the quantity of people living in extreme poverty and helps compare poverty levels between nations. As the cost of living increases, poverty lines increase as well. Since 1990, the international poverty line rose from $1 per day, to $1.25 per day, and most as of late in 2015 to $1.90. This implies that $1.90 is important to purchase what $1 could in 1990. In addition to the lowest-income poverty line at $1.90, the World Bank additionally reports poverty rates utilizing two new international poverty lines: a lower middle-income line set at $3.20 per day and an upper middle-income line set at $5.50 per day.

Assessments from the 2017 Myanmar Living Conditions Survey (MLCS) discover that 24.8 percent of the population is poor. The poverty line in 2017 was 1,590 kyats per adult equal every day. The consumption levels at or below 1,590 kyats for each day is viewed as poor. The poverty headcount is 2.7 times lower in urban areas (11.3 percent) than in rural areas (30.2 percent).The quantity of poor people is 6.7 times higher in rural areas than in urban areas. Poverty rates vary significantly by state and region: poverty is the most prevalent in Chin State, where right around six out of ten people are poor, and in Rakhine State, where around four out of ten are poor. At the opposite site, Tanintharyi, Mandalay, and Yangon Regions have the lowest poverty rates, in all cases somewhere in the range of 13 and 14 percent.

In 2017, poverty rate at national poverty line for Myanmar was 24.8 percent and poverty rate at national poverty line of Myanmar fell gradually from 48.2 percent in 2005 to 24.8 percent in 2017. The decrease in poverty is noticeable in both rural and urban areas, but has been quicker in urban areas. Poverty depth and seriousness have also decreased significantly since 2005. A third of the population is highly vulnerable to falling into poverty in the later on, despite not being poor in 2017. As the poverty rate declined since 2005, the portion of the population that is non-poor has expanded, yet the portion of non-poor named as helpless has additionally developed by 18 percent. By and large, poor family units have very nearly multiple times a greater number of kids than non-poor households. Poor families are more likely to have a head with

practically no education. Households that own land or have expanded from agricultural work have government assistance.

2.1.5 Food Poverty

Food poverty can be defined as the inability to obtain healthy affordable food. This may be because people lack shops in that area or have trouble reaching it. Other factors influencing food access are the availability of a range of healthy goods in local shops, income, transport, fear of crime, knowledge about what constitutes a healthy diet, and the skills to create healthy meals. Food poverty is important to guarantee that individuals can get a healthy diet in a socially worthy manner and have adequate conviction about a healthy diet will make sure for families.

Food poverty can affect anyone who does not have private transport, who feels it cannot be afford the money to buy healthy food, or who does not know how to cook healthy food. Many people living on low incomes, with poor access to transport facilities or a lack of choice for shopping, continue to consume a healthy diet despite the obstacles set in the path, while others struggle. So, whilst it is true that people on low incomes are more likely to experience food access problems it is not true that as a result who are always eat a poor diet. However, many do and this directly effects of the poorer's health and life chances.

2.2 Empirical Review

Anyanwu (2013) conducted "Determining the Correlates of Poverty for Inclusive Growth in Africa". The study is an attempt to contribute to the design of inclusive growth policies. It examines the correlates of poverty headcount index of international poverty line at US$1.25 per day with multivariate models using data on 43 African countries. The study estimates, using available cross-sectional time series data over the period, 1980 and 2011 for 43 countries suggest that, higher levels of income inequality, primary education alone, mineral rents, inflation, and higher level of population tend to increase poverty in Africa and therefore bad for poverty reduction and inclusive growth in the continent. The majority finding is that inequality fuels poverty in African countries, policy makers need to tackle this challenge head-on.

Ahmad (2015) studied "Economic Wealth Index: A Tool to Study the Economic Health of Districts of Selected States in India" by using Principal Component Analysis.

The objectives are to create the economic wealth index using Principal Component Analysis, to classify Economic Wealth Index values into categories, to rank the districts of the selected States and to compare the Economic Wealth Index values of the districts of Maharashtra with Human Development Index. Principal Component Analysis Method (PCA) is used to construct the Economic Wealth Index based on quality data, condition of houses and presence of various basic facilities like drinking water, electricity and toilet given by Census (2001). The results point out that correlation between composite economic wealth index (CEWI) and mean number of births is significant only in the pooled data, the State of Maharashtra and Jharkhand.

Azpitarte (2010) examined "Measuring Poverty Using Both Income and Wealth: An Empirical Comparison of Multidimensional Approaches Using Data for the US and Spain." The author presents a comparative analysis of the approaches to poverty based on income and wealth that have been proposed in the literature. Two types of approaches are considered: that look at income and wealth separately when defining the poverty frontier, and the two dimensions are integrated into a single index of welfare. The finding is that the incidence of poverty in the two countries varies significantly depending on the poverty definition adopted. Despite the variation, the results suggest that the poverty problem is robust to changes in the way poverty is measured. Regarding the identification of the poor, there is a high level of misclassification between the poverty indices: for most of the pairwise comparisons, the proportion of households that are misclassified is above 50 percent. Interestingly, the rate of misclassification in the United States is significantly lower than in Spain. So that the higher correlation between income and wealth in the United States contributes to explaining the greater overlap between poverty indices in the country.

Block, et al., (2010) studied "Child Undernutrition, Household Poverty and National Income in Developing Countries". The paper reveal that the impact of income in reducing poverty is similar across declines of the poverty distribution, the effect of income in decreasing underweight varies significantly and systematically across declines of the underweight distribution. The main result, using quantile regression, is that poverty and underweight rates both have similar income elasticity when the incidence is high, however when it is low the income elasticity of child underweight becomes much smaller than the income elasticity of household poverty.

Chani, et al., (2011) studied "Poverty, Inflation and Economic Growth: Empirical Evidence from Pakistan". The study attempts to investigate the part of financial development (economic growth) and inflation in explaining the prevalence of poverty in Pakistan. The creators considered the determinants of poverty in Pakistan between 1972 and 2008 using time series and find that inflation has positive effect on poverty, economic growth has negative effect on poverty; and the impacts of investment and trade openness in poverty reduction are not significant.

Chasekwa, et al., (2018) studied "Measuring Wealth in Rural Communities: Lessons from the Sanitation, Hygiene, Infant Nutrition Efficacy (SHINE) Trial" by using principal component analysis. This paper developed and validated a household wealth index using baseline data for the Sanitation, Hygiene, Infant Nutrition Efficacy trial conducted in rural Zimbabwe between 2012 and 2017. This study found that a "standard" approach (principal components analysis) using a rich, but still relatively parsimonious set of variables is strongly associated with a wide range of indicators of wealth and is both internally and externally valid. In addition to providing evidence of the validity of the index, the paper provides a template for others constructing such indices, including a method for placing smaller regional samples into the broader context of a country when national survey data are available.

Carter and Barrett (2006) conducted "The Economics of Poverty Traps and Persistent Poverty". The objective of the paper is to outline asset-based approach by developing the conceptual foundations for asset-based poverty measures to permit a forward-looking approach and help distinguish and ultimately understand the structure and determination of poverty. Longitudinal or panel data and assets-based approach are used in the journal. The results show that only a small fraction (15 per cent) clearly fell into poverty for stochastic reasons, while completely 51 per cent of the people who fell behind suffered asset losses left fundamentally poor in the latter survey period.

Douidich, et al., (2013) studied "Estimating Quarterly Poverty Rates Using Labor Force Surveys". In the study, Cross-survey imputation techniques have been used to estimate poverty rates for Morocco and obtained a quarterly series for the period 2001- 2010. Standard imputation approach is also used in the case of missing data. The result shows that poverty in Morocco has continued the decline beyond the 2007-2008, global financial crisis, which may be explained in part by the positive performance of the agricultural sector.

Dekker (2006) studied “Estimating Wealth Effects without Expenditure Data: Evidence from rural Ethiopia” by using both factor analysis and principal component analysis. This study found that households with a higher economic status experience significantly fewer weeks of food insecurity compared to households with a lower economic status. Moreover, the results suggest the relation between household economic status and food security is measured at least as precisely when an assets-based wealth index are compared to a wealth proxy defined as per capita consumption.

McKernan and Ratcliffe (2002) studied “Transition Events in the Dynamics of Poverty”. The study mainly used descriptive analyses and multivariate analyses. Descriptive analyses of poverty entries and exits show that shifts in household structure (i.e., transitions from a two-adult to a female-headed household and vice versa) are relatively rare events in the population, but individuals who experience the events are the most likely to transition into or out of poverty. The main finding from the multivariate analyses is that changes in employment, not household composition, are the most strongly related to poverty transitions and, is a new finding in that earlier studies have not examined the relationship between household events and poverty in a multivariate framework.

Mohanty (2009) explored “Alternative Wealth Index and Health Estimates in India at International Institute for Population Sciences” by using PCA to construct a wealth index or to measure socio-economic status has been well established and validated. This study found that rural and urban areas should be debated and also theoretical and statistical significance in the construction of a wealth index should be examined. The results of study suggest the construction of a separate wealth index for rural and urban areas. Although the value if consumer durables varies considerably across countries it is useful to try and derive information about the quality and market prices of consumer durables in constructing a more context-specific and robust wealth index.

Nussbaumer, et al., (2012) explored “Measuring Energy Poverty: Focusing on What Matters.” The purposes of the report was to discusses the adequacy and applicability of existing instruments to measure energy poverty and proposes a new composite index to measure energy poverty . The methodology utilizing in this report are Oxford Poverty and Human Development Initiative (OPHI) and The Multidimensional Energy Poverty Index (MEPI): a new metric to measure and report

on energy poverty. The main findings of the report shows a stark contrast with regard to the level of energy poverty between the capital, where the MEPI is similar to that of the country of Morocco, and the Western and North Eastern districts which suffer from severe energy poverty.

Rutstein and Staveteig (2014) presented “Making the Demographic and Health Surveys Wealth Index Comparable”. This paper aims to present inequality measures for a set of demographic and health indicators, to determines the contribution of relative and absolute measures of poverty to progress in the indicators and to describes a newly developed methodology for calculating wealth indexes comparable across country and time. There is several technique to calculate the Comparative Wealth Index: comparison with a baseline, use poverty levels and other common items that are in each demographic and health surveys as “anchoring” points. The result shows that Comparative Wealth Index produces in terms of ranking of countries and regions that generally comport with per capita income measures and may help to sort out the effects due to health programs focused on the poor versus the effects due to changes in economic status of the population.

Schmitt (2013) examined “A Regional Perspective on Poverty in Myanmar”. The purpose of the report is to attempt a regional analysis of the causes of poverty. Based on the analysis, potentials to improve the wellbeing of the regional populations are outlined. Intensive analysis was used in the report. The report looks predominantly at expenditure poverty. It focuses on the median poor (versus the average), and takes as "unit of measurement" the median poverty gap. The key reasons of why people enter into poverty are the effects of storms, floods and stagnant water, and the lack of work. The main findings of report is that poor households offer more labor than non-poor States and Regions (S/R) of the Union. The percentage of household members working on own account isn't related with the household expenses. Myanmar loses on average yearly almost 8 percent of the harvest due to natural calamities (storm, flood, drought, pests). As there are numerous explanations behind poverty, so does Myanmar have many options to lift people out of poverty. As the poor are less productive, and the existing productive capacity is also used to a lower degree, poverty hampers the current and especially long term productive potential of the nation. It is consequently in the helpful as well as financial interest to reduce poverty.

Lustig (2017) explored “Fiscal Policy, Income Redistribution, and Poverty Reduction in Low- and Middle-Income Countries”. The study uses comparative fiscal incidence analysis and the main objective here is to analyze the impact of fiscal policy on inequality and poverty in twenty-nine low- and middle-income countries from around 2010. The results show the decrease in imbalance initiated by the money part of the financial framework in the twenty-nine countries analyzed here is very heterogeneous. While the results regarding the favorable to need of spending on education and health are quite encouraging, a proviso is altogether. Guaranteeing access to and encouraging utilization of public education and health services for the poor isn't sufficient. The quality of schooling and healthcare provided by the government is low.

The Government of Myanmar and the World Bank (2017) presented “An Aalysis of Poverty in Myanmar”. The important target of the survey was to give refreshed data on living conditions and socio-economic indicators in the nation, mainly using joint analysis and the key goal was to give an assessment of poverty in 2015 using data from the Myanmar Poverty and Living Conditions Survey. The report finds that living standards have improved in Myanmar between 2004/05 and 2015, poverty has declined from 32.1 percent in 2004/05 to 19.4 percent in 2015, average real expenditure grew, durable goods ownership increased and households saw an expansion of the dietary base.

Hayati, et al., (2004) studied “Combining Qualitative and Quantitative Methods in the Measurement of Rural Poverty”. The aim of the research was constructing a suitable index for measuring poverty at household level in rural areas of Iran. The study uses triangulation survey and participatory rural appraisal to research poverty in Fars province in Iran. An extreme purposive sampling technique was used to select four villages. Two of the villages were relatively developed and two relatively less developed. The principle discoveries are that financial degrees of imbalance are high. Fundamental necessities are regularly unfulfilled; food consumption might be beneath least calorific prerequisite, salaries are frequently too low to satisfy basic food and non-food needs; access to health facilities, schooling, minimum housing and clothing, safe drinking water and sanitation is regularly lacking.

Zewotir, et al., (2015) studied “Analysis of Demographic and Health Survey to Measure Poverty of Household in Rwanda” at University of KwaZulu-Natal by using principal component analysis. This study revealed that the demographic and spatial

profiles of poor households in Rwanda are: education of household head, gender of household head, age of the household head, place of residence, Region and size of the household. This study found that the majority of poor households have low standards of education. Therefore, there is a need to improve existing access to higher education in order to speed up poverty reduction. The Rwanda data analysis was the significant predictors of poverty of the household in Rwanda.

CHAPTER 3
METHODOLOGY

This chapter provides the analytical methods used for the study on determinants of poverty and socio-economic data. The methods are multiple regression, binary logistic regression and wealth index by using principal component analysis (PCA).

3.1 Regression Analysis and Regression Model

The earliest form of regression was the method of least squares, which was published by Legendre in 1805, and by Gauss in 1809. The English Scientist Sir Francis Galton (1822-1911) was first applied the term regression to statistics. Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables. The dependent variable Y is also referred to response variable, predicted variable or output variable while the independent variable X is called the predictor, the regressor or the explanatory variable.

Regression techniques are one of the most popular statistical techniques used for predictive modeling and data mining tasks. On average, analytics professionals know only 2-3 types if regression which is commonly used in real world that are linear and logistic regression algorithms designed for various types of analysis.

Regression analysis is primarily used for two conceptually distinct purposes. First, regression analysis is widely used for prediction and forecasting, where its use has substantial overlap with the field of machine learning. Second, in some situations regression analysis can be used for infer causal relationships between the independent and dependent variables. Importantly, regression by themselves only revel relationships between a dependent variable and a correlation of independent variables in a fixed dataset. To use regression for prediction or to infer causal relationships, respectively, a researcher must carefully justify why existing relationships have predictive power for a new context or why a causal interpretation. The latter is especially important when a researcher hopes to estimate causal relationships using observational data.

In practice, a researcher first selects a model to estimate and then uses chosen method to estimate the parameters of that model. Regression models involve the following components:

(i) The unknown parameters often denoted as a scalar or vector β.

(ii) The independent variables, which are observed in data and are often denoted as a vector X_i (where i denote a row of data).

(iii) The dependent variable, which are observed in data and often denoted using the scalarY_i .

(iv) The error terms, which are not directly observed in data and are often using the scalar e_i .

In various fields of application, different terminologies are used in place of dependent and independent variables.

Most regression models propose that is a function of and, with representing an additive error term that may stand in for un-modeled determinants of or random statistical noise:

$$Y_i = f(X_{i`,}\beta) + e_i \quad (3.1)$$

When Y depends on only one X in simple regression analysis, but when Y depends on more than one independent variable in multiple regression analysis. If the relation between the dependent and the independent variables is linear, then that is linear regression analysis.

3.2 Types of Regression

One of the most popular statistical techniques used for predictive modeling and data mining tasks is regression techniques. There are many types of regression which are linear regression(i) simple linear regression (only one independent variable) and (ii) multiple linear regression (more than one independent variable), polynomial regression, logistic regression, quantile regression, ridge regression, lasso regression, elastic net regression, principal components regression (PCR), partial least squares (PLS) regression, support vector regression, ordinal regression, Poisson regression, negative binomial regression, quasi Poisson regression, cox regression and Tobit regression.

3.2.1 The Simple Linear Regression Model

Simple linear regression is used to model the relationship between two continuous variables. Often, the objective is to predict the value of an output variable 16 (or response) based on the value of an input (or predictor) variable. The simple linear regression model is defined as

$$Y_i = \beta_{0+}\beta_1 X_{i+}\varepsilon_i \quad (3.2)$$

Where,

Y_i = dependent variable or response variable for observation i.

X_i = independent variable or explanatory variable for observation i.

β_0 = intercept (or) regression constant

β_1 = slope (or) regression coefficient

ε_i = error term (or) residual

The above equation is called the simple linear regression model and is also called the two-variable linear regression model or bivariate linear regression model because it relates the two variable X and Y.

3.2.2 Assumption Underlying the Simple Regression Model

Firstly, the simple linear regression requires normality that error around the line of regression be normally distributed at each value of X. Like the t test and the ANOVA F test, regression analysis is fairly robust against departures from the normality assumption. As long as the distribution of the error around the line of regression at each level of X is not extremely different from a normal distribution.

Secondly, it requires homoscedasticity that the variation around the line of regression be constant for all values of X. This means that the errors vary the same amount when X is a low value as when X is a high value. The homoscedasticity assumption is important for using the least squares method of determining regression coefficients. If there are serious departures from this assumption, either data transformations or weighted least- squares methods can be applied.

Thirdly, it requires independence that the errors be independent for each value of X. This assumption is particularly important when data are collected over a period of time. In such situations, the errors for a particular time period are often correlated with the data that collected from previous time period.

3.3 The Multiple Linear Regression Model

The multiple linear regression models were developed by Karl Pearson in 1908. Multiple regression is an extension of simple linear regression. It is used to predict the value of a variable based on the value of two or more other variables. It is necessary to determine the equation for the average relationship between the variable. In the general case, using k refers to the number of explanatory variables. The equation which

describes how the dependent variable y is related to the independent variables and an error term ε_i is called the multiple regression model. Multiple regression models take the following form.

$$Y_i = \beta_0 + \beta_1 X_{i1} + +\beta_2 X_{i2} + \cdots + +\beta_k X_{ik} + \varepsilon_i \quad (3.3)$$

Where, Y_i = value of the dependent variable in the i^{th} trial, of observation

β_0 = constant in the regression equation, which indicates the value of Y when all $X_k = 0$

$\beta_1, \ldots, \beta_k$ = regression coefficients associated with each of the X_k independent variable

X_{ij} = value of the j^{th} independent variable in the i^{th} trial, or observation, associated with the process of sampling.

ε_i = the random error in the i^{th} trial or observation, associate with the process of sampling.

3.3.1 Assumption Underlying the Multiple Linear Regression Models

Firstly, the multiple linear regression requires that the relationship between the independent and dependent variable to be linear. The linearity assumption can be tested with scatter diagram.

Secondly, it requires that the errors between observed and predicted values (i.e. the residuals of the regression) should be normally distributed. The assumption may be checked by looking at a histogram or P.P. plot.

Thirdly, multiple linear regressions assumed that there is no multicollinearity in the data. Multicollinearity occurs when the independent variable are too highly correlated with each other.

The last assumption of multiple linear regressions is homoscedasticity. A scatter plot of residuals versus predicted values is good way to check for homoscedasticity. If there is a cone-shaped pattern, there should be no clear pattern in the distribution and the data is heteroscedasticity.

3.3.2 Multicollinearity

Multicollinearity refers to the case in which two or more independence variables in the regression model are highly correlated, making it difficult or impossible to isolate

the individual effects on the dependent variable. With multicollinearity, the estimated OLS coefficients may be statistically insignificant even though may be "high". Detecting multicollinearity may be testes with three central criteria

Correlation matrix when computing the matrix of Pearson's Bivariate Correlation among all independent variables, the correlation coefficient r is greater than 0.7. If correlation is less than 0.5, it can be correlated that multicollinearity is not problem.

The tolerance measures the influence of one independence variable on all other independent variables; the tolerance is calculated with an initial linear regression analysis. Tolerance is defined as T= 1-R^2 for this first step regression analysis. With T<0.1 there might be multicollinearity in the data.

Variance Inflation Factor (VIF) of the linear regression is defined VIF$\frac{1}{T}$. With VIF>10 there is an indication that multicollinearity may be present.

3.3.3 Homoscedasticity and Heteroscedasticity

A critical assumption of the classical linear regression model is that the disturbances u_i have all the same variance. When this condition holds, the error terms are homoscedastic, which means the error have the same scatter regardless of the value of X. When the scatter of the errors is different, varying depending on the value of one or more of the independence variables, the error terms are heteroscedastic.

The opposite of homoscedasticity is the phenomenon of the heteroscedasticity, where the error term can be formulated as a function of x_i. This can be described in mathematical terms as Var ($^{e_i}/_{x_i}$) =σ where e_i is the error term and x_i is the measure of some covariate.

3.3.4 The Coefficient of Multiple Determination (R^2)

The coefficient of multiple determinations (R^2) measures the proportion of variation in the dependent variable that can be predicted from the set of independent variables in a multiple regression equation. When the regression equation fits the data well, R^2 will be large (i.e, close to 1); and vice versa.

The coefficient of multiple correlations can be defined in terms of sums of squares:

SSR=$\sum(\hat{Y}_1\text{-}\bar{Y})^2$

$$SSE=\sum(Y_i-\hat{Y}_i)^2$$

$$SST=\sum(Y_i-\bar{Y})$$

$$R^2 = \frac{SSR}{SST} \quad (3.4)$$

(OR)

$$R^2=1-\frac{SSE}{SST} \quad (3.5)$$

Where, SSR is the sum of squares due to regression, SST is the total sum of squares, ŷ is the predicted value of the dependent variable, y is the dependent variable mean, and y is the dependent variable raw score.

3.3.5 The Adjusted Coefficient of Multiple Determination (R_{adj}^2)

Adjusted coefficient of determination is the adjusted value of the coefficient of determination in which the number of variables of the data set is taken into consideration. It determines the fitting of the multiple regression equations for the sample data. It is also known as adjusted $\bar{R}^2$ or $\bar{R}^2$. It determines the proportion of variation which is determined by the estimated line of regression.

The adjusted coefficient of determination takes the values between 0 and 1. It explains the percentage of variation of the independent variables that affect the dependent variables. If the adjusted coefficient of determination is closer to 1, it indicates that the estimated equation of regression fits the data. The value of the adjusted coefficient of determination increases only when new independent variables are added which in turn increases the power of the regression equation. Hence, adjusted coefficient of determination is widely used in multiple regression analysis. In this way, it is very useful measure than the coefficient of determination.

3.3.6 Test of the Overall Significance of a Multiple Regression Model

In general, an F-test in regression compares the fits of different linear models. Unlike t-tests that can assess only one regression coefficient at a time, the F-test can assess multiple coefficients simultaneously. The F-test of the overall significance is a specific form of the F-test. It compares a model with no predictors to the model that you specify. The overall F-test is used to test for the significance of overall multiple regression models. The ANOVA producer tests the null hypothesis that all the β values are zero against the alternative that at least one β is not zero.

Table 3.1 ANOVA Table for Multiple Regression Analysis

Source of Variation	Sum of Squares	Degree of Freedom	Mean Square	F
Regression	SSR=$\sum(\hat{Y}_1\text{-}\bar{Y})^2$	k	MSR = $\frac{SSR}{k}$	MSR/MSE
Residual	SSE=$\sum(Y_i\text{-}\hat{Y}_i)^2$	n-k-1	MSE = $\frac{SSE}{n-k-1}$	
Total	SST=$\sum(Y_i\text{-}\bar{Y})^2$	n-1		

$$F = \frac{\text{MSR}}{\text{MSE}} \tag{3.6}$$

SSR= Regression Sum of Squares

SSE= Error Sum of Squares

SST=Total Sum of Squares

k = the number of independent variable in the regression model

n-k-1 = the degrees of freedom for residual

MSE = the mean squares of error

MSR = Mean Square Error

If F$\geq F_{\alpha,k,n-k-1}$, reject H_0, otherwise accept H_0, where $F_{\alpha,k,n-k-1}$ is the critical F value at the α level of significance and k numerator df and (n-k-1) denominator df.

If the null hypothesis is rejected, it can be concluded that one or more of the parameters in the model is not equal to zero. Thus, the overall relationship between the dependent variable Y and the independent variables $X_1, X_2, \ldots, X_k$ is significant. However, if the null hypothesis is accepted, it can be concluded that there is an overall significant relationship and estimated regression equation cannot explain the variation in the dependent variable.

3.3.7 Estimated Multiple Regression Equation

If the values of $\beta_0, \beta_1, \ldots, \beta_k$ are known, the previous equation is used to calculate the mean of value of y at the given values of $X_1, X_2, \ldots, X_k$. In general, the parameter values will not know and will have to estimate them from sample data. Using this sample, an estimated multiple regression equation can develop which takes the following form

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \cdots + b_kX_k \tag{3.7}$$

Where, $b_0, b_1, \ldots, b_k$ is the estimated value of the parameters $\beta_0, \beta_1, \ldots, \beta_k$ and $\hat{y}$ is the estimated value of the dependent variable. The estimation procedure for multiple

regression is nearly identical to simple regression. The least squares method is used to come up with "best" fit.

3.3.8 Inference of Testing for Significance

The significance tests for the simple regression model were the *t* test and the *F* test and always generated the same conclusion. If the null was rejected, concluded that $\beta_1 \neq 0$. In multiple regression, the t test and the F test have different purposes.

(i) The F test is used to determine whether there exists a significant relationship between the dependent variable and the entire set of independent variables in the model; thus the F test is a test of the overall significance of the regression.

(ii) If the F test shows that the regression has overall significance, the t test is then used to determine whether each of the individual independent variables is significant. A separate t test is used for each of the independent variables; thus the t test is a test for individual significance.

F-test

The multiple regression model is defined as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u \qquad (3.8)$$

The hypothesis for the F test takes the following form

Null Hypothesis : $\beta_0 = \beta_1 = \beta_2 = \cdots = \beta_k = 0$

Alternative Hypothesis: At least one $\beta_i \neq 0$

If the null is rejected, to conclude that one or more of the parameters in the model is not equal to zero. Thus, the overall relationship between the dependent variable *y* and the independent variables $x_1, x_2 ..., x_k$ is significant. However, if the null is not rejected, to conclude that there is an overall significant relationship and regression does not significantly to explain the variation in the dependent variable.

The test statistic for the *F* test is

$$F = \frac{MSR}{MSE}$$

Where, the MSR is the mean square due to the regression which is equal to

$$MSR = \frac{SSR}{K}$$

And, the MSE is the mean square error which is equal to

$$MSE = \frac{SSE}{n-k-1}$$

Where, n-k-1 is the degrees of freedom and K are the number of independent variables. The decision rule for the F-test takes the following form

Reject the null hypothesis if : $F > F_{\alpha,k,n-k-1}$

Do not reject null hypothesis if: $F \leq F_{\alpha,k,n-k-1}$

Where, $F_{\alpha,k,n-k-1}$ is based on the *F* distribution with *K* degrees of freedom in the numerator, n-k-1 degrees of freedom in the denominator, and a probability of α in the upper-tail of the probability distribution.

t-test

A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of two groups, which may be related in certain features. The t-test is one of all tests used for the purpose of hypothesis testing in statistics. Calculating a t-test requires three key data values which include the difference between the mean values from each data set (called the mean difference), the standard deviation of each group, and the number of data values of each group. There are several different types of t-test that can be performed depending on the data and type of analysis required.

The t test of significance works as it did for simple regression models. For any parameter β_i the hypotheses take the form;

Null Hypothesis : $\beta_i = 0$

Alternative Hypothesis : $\beta_i \neq 0$

The t-test statistic for $\hat{\beta}_i$ is simple to compute given $\hat{\beta}_i$ and the standard error:

$$t = \frac{\hat{\beta}_i}{Se(\hat{\beta}_i)}$$

The decision rule for this test takes the following form

Reject the Null Hypothesis if : $t < -t_{\alpha/2,n-k-1}$ or $t > t_{\alpha/2,n-k-1}$

Do not reject the Null Hypothesis if : $-t_{\alpha/2,n-k-1} \leq t \leq t_{\alpha/2,n-k-1}$

Coefficient of Determination

In the simple linear regression, the total sum of squares, the total variation in the dependent variable (SST), can be broken into two parts: the sum of squares due to regression (SSR) and the sum of squares due to error (SSE). This same partition works for multiple regression.

$$SST = SSR + SSE \quad (3.9)$$

The quality of the fit for the regression can be calculated by computing the coefficient of determination. The coefficient of determination is still computed as

$$R^2 = \frac{SSR}{SST} \tag{3.10}$$

3.4 Logistic Regression

The logistic function was developed as a model of population growth and named "logistic" by Pierre Francois Verhulst in the 1830s and 1840s, under the guidance of Adolphe Quetelet; see Logistic function History for details. In earliest paper (1838), Verhulst did not specify how he fit the curves to the data. In more detailed paper (1845), Verhulst determined the three parameters of the model by making the curve pass through three observed points, which yielded poor predictions. Sir David Roxbee Cox (1958) is a prominent British Statistician that addressed the case of logistic regression. A detailed history of the logistic regression is given in Cramer (2002).

In statistics, the logistic model (or logit model) is used to model the probability of a certain class or event existing such as pass and fail, win and lose, alive and dead or healthy and sick. Each object being detected in the image would be assigned a probability between 0 and 1, with a sum of one. Logistic regression is a statistical model that in its basic form uses a logistic function to model a binary dependent variable, although many more complex extension exist. In regression analysis, logistic regression (or logit model) is estimating the parameters of a logistic model (a form of binary regression). Normally, 1 is coded to positive event such as presences, yes, use, success, and so on. The probability of success is denoted as P while for failure denoted as (1-p). This types of variable is called Bernoulli (or binary variable). The independent variables are often called as covariates.

There are three types of logistic regression. They are binary or binomial logistic regression, multinomial logistic regression and ordinal logistic regression. In these three logistic regression, binary or binomial logistic regression are used in this study.

3.4.1 Assumptions of Logistic Regression

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

Firstly, it does not need a linear relationship between the dependent and independent variable. Logistic regression can handle all sorts of relationships because, it applies a non-linear log transformation to predict odd ratio. Secondly, the independent variables do not need to be multivariate normal-although multivariate normality yield a more stable solution. Also the error terms (the residuals) do not need to be multivariate normality distributed. Thirdly, homoscedasticity is not needed. Logistic regression does not need variances to be heteroscedastic for each level of the independent variables. Lastly, it can handle ordinal and nominal data as independent variables. The independent variables do not need to be metric (interval or ratio scaled).

However some other assumptions still apply Binary logistic regression requires the dependent variable to be binary and ordinal logistic to regression requires the dependent variable to be ordinal. It can be found that reducing an ordinal or even metric variable to dichotomous level ordinal logistic regression in these cases. Secondly, since logistic regression assume that P(Y=1) is the probability of the event occurring, it is necessary that the dependent variable is coded accordingly. Thirdly, the model should be fitted correctly. Neither over fitting nor under fitting should occur. That is only the meaningful variables should be included, but also all meaningful variables should be included. A good approach to ensure this is to use a stepwise method to estimate the logistic regression.

Fourthly, the error terms need to be independent. Logistic regression requires each observation to be independent. That is the data point should not be from any dependent sample design, e.g. before and after measurements, or matched pairings. Also the model should have little or no multicollinearity. That is the independent variables should be independent from each other. However, there is the option to include interaction effects of categorical variables in the analysis and the model. If multicollinearity is present centering the variables might resolve the issue, i.e. deducting the mean of each variable. If this does not lower the multicollinearity, a factor analysis with orthogonally rotated factors should be done before the logistic regression is estimated.

Fifthly, logistic regression assume linearity of independent variables and log odds, whilst it does not requires the dependent and independent variables to be related linearly, it requires that the independent variables are linearly related to the log-odds. Otherwise the test un-estimates the strength of the relationship and reject the

relationship too easily, that is being not significant (not rejecting the null hypothesis) where it should be significant. A solution to this problem is the categorization of the independent variables. That is transformation metric variables to ordinal level and then including them in the model. Another approach would be to use discriminant analysis, if the assumptions of homoscedasticity, multivariate normality, and absence of multicollinearity are met.

Lastly, it requires quite large sample sizes. Because maximum likelihood (ML) estimates are less powerful than ordinary least squares (OLS) such as simple linear regression, multiple linear regression. While OLS needs 5 cases per independent variable in the analysis, ML needs at least 10 cases per independent variable and some statisticians recommended at least 30 cases for each parameter to be estimated.

3.5 The Model of Binary Logistic Regression

The dependent variable in logistic regression is usually binary that is the dependent variable can take the value 1 with a probability of success p, or the value 1 with a probability of failure 1-p. This type of variable is called a binary variable.

Binary logistic regression modeling can be used in many situations to answer research questions and to predict the odds of being a case based on the values of the independent variables (predictors). This regression models is the relationship between a set of predictors and a binary response variable. A binary response has only two possible values, such as win and loses. Use a binary regression model to understand how changes in the predictor values are associated with changes in the probability of an event occurring. The model can be written as:

$$\text{Logit}\,(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon \qquad (3.11)$$

3.6 Principal Component Analysis

Principal component analysis (PCA) was invented in 1901 by Karl Pearson**,** it is the process of computing the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest.

PCA is used in exploratory data analysis and for making predictive models. It is commonly used for dimensionality reduction by projecting each data point onto only the first few principal components to obtain lower-dimensional data while preserving

as much of the data's variation as possible. The first principal component can equivalently be defined as a direction that maximizes the variance of the projected data. The principal component can be taken as a direction orthogonal to the first principal components that maximizes the variance of the projected data.

From either objective, it can be shown that the principal components are eigenvectors of the data's covariance matrix. Thus, the principal components are often computed by eigen-decomposition of the data covariance matrix or singular value decomposition of the data matrix. PCA is the simplest of the true eigenvector-based multivariate analyses and is closely related to factor analysis. Factor analysis typically incorporates more domain specific assumptions about the underlying structure and solves eigenvectors of a slightly different matrix. PCA is also related to canonical correlation analysis (CCA). CCA defines coordinate systems that optimally describe the cross-covariance between two datasets while PCA defines a new orthogonal coordinate system that optimally describes variance in a single dataset.

A PCA is a statistical procedure concerned with elucidating the covariance structure of a set of variables. It is a method that projects a dataset to a new coordinate system by determining the eigenvectors and eigenvalues of a matrix. It involves a calculation of a covariance matrix of a dataset to minimize the redundancy and maximize the variance.

Given a data matrix with p variables and n samples, the data are first centered on the means of each variable. This will insure that the cloud of data is centered on the origin of our principal components, but does not affect the spatial relationships of the data nor the variances along the variables. The first principal components (Y_1) is given by the linear combination of the variables $X_1, X_2, \dots, X_p$

$$Y_1 = a_{11}X_1 + a_{12}\,X_2 + \dots + a_{1p}X_p \tag{3.12}$$

Or, in matrix notation

$$Y_1 = a^T{}_1 X$$

The first principal component is calculated such that it accounts for the greatest possible variance in the data set. Of course, one could make the variance oY_1 f as large as possible by choosing large values for the weights, $a_{11}, a_{12} \dots . a_{1p}$. To prevent this, weights are calculated with the constraint that the sum of squares is 1.

$$a_{11}{}^2 + a_{12}{}^2 + \dots + a_{1p}{}^2 = 1 \tag{3.13}$$

The second principal component is calculated in the same way, with the condition that it is uncorrelated with (i.e. perpendicular to) the first principal component and that it accounts for the next highest variance.

$$Y_1 = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p} \quad (3.14)$$

This continues until a total of p principal components have been calculated, equal to the original number of variables. The sum of the variances of all of the principal components will equal the sum of the variances of all of the variables, that is, all of the original information has been explained or accounted for. Collectively, all of the transformations of the original variables to the principal components is

$$Y = XA$$

Calculating the transformations or weights requires a computer for all but the smallest matrices. The rows of matrix A are called the eigenvectors of matrix Sx, the variance-covariance matrix of the original data. The elements of an eigenvector are the weights a_{ij} and are also known as loadings. The elements in the diagonal of matrix Sy, the variance-covariance matrix of the principal components, are known as the eigenvalues. Eigenvalues are the variance explained by each principal component, and to repeat, are constrained to decrease monotonically from the first principal component to the last. The eigenvalues are commonly plotted on a scree plot to show the decreasing rate at which variance is explained by additional principal components.

The positions of each observation in this new coordinate system of principal components are called scores and are calculated as linear combinations of the original variables and the weights, a_{ij}. For example, the score for the r^{th} sample on the k^{th} principal component is calculated as

$$Y_{rk} = a_{1k}x_{r1} + a_{2k}x_{r2} + \ldots + a_{pk}x_{rp} \quad (3.15)$$

In interpreting the principal components, it is often useful to know the correlations of the original variables with the principal components.

The correlation of variable X_i and principal component Y_j is

$$r_{ij} = \sqrt{a_{ij}^2 Var(Y_j) / S_{ii}} \quad (3.16)$$

Because reduction of dimensionality that is focusing on a few principal components versus many variables is a goal of principal components analysis, several criteria have been proposed for determining how many PCs should be investigated and how many should be ignored. One common criteria is to ignore principal components

at the point at which the next PC offers little increase in the total variance explained. A second criteria is to include all the PCs up to a predetermined total percent variance explained, such as 90 percent. A third standard is to ignore components whose variance explained is less than 1 when a correlation matrix is used or less than the average variance explained when a covariance matrix is used, with the idea being that such a PC offers less than one variable's worth of information. A fourth standard is to ignore the last PCs whose variance explained is all roughly equal.

Principal components are equivalent to major axis regressions. As such, principal components analysis is subject to the same restrictions as regression, in particular multivariate normality. The distributions of each variable should be checked for normality and transforms used where necessary to correct high degrees of skewness in particular. Outliers should be removed from the data set as dominate the results of a principal components analysis.

3.7 Wealth Index

The wealth index is a composite measure of the cumulative living standard of a household. It is calculated using data on a household's ownership of selected set of assets, such as televisions, bicycles, and cars, dwelling characteristics such as flooring material, type of drinking water source, and toilet and sanitation facilities. The Wealth Index considers characteristics that are related to wealth status, avoiding variables that do not represent an asset, or outcome variables. Each household asset for which information is collected is assigned a weight or factor score generated principal components analysis. The resulting asset scores are standardized in relation to a standard normal distribution with a mean of zero and a standard deviation of one.

Wealth is a household characteristic that often has a large effect on health. The wealth index allows for the identification of poor, health care and particularly to the wealthy. Wealth index allows to identify how households' economic status affects health outcomes. Wealth is a major supporter for optimum health because it gives the financial freedom to make health choices.

Asset based wealth indices are widely used instruments for measuring the economic situation of households in developing countries. Most household surveys currently available for the countries include such an index based on the possession of consumer durables and housing characteristics. Wealth indices owe this success to the

intuitive appeal, wide availability, ease of computation, and reliability of measurement. However, in spite of the desirable properties, households will suffer of one great problem and not comparable between surveys. So, human's occupation, income and wealth are related to the socio-economic composition.

Wealth is the value of all natural, physical and financial assets of owned by a household, reduced by liabilities. Wealth, a set of economic reserves or assets, presents a source of security providing a measure of a household's ability to meet emergencies, absorb economic shocks, or provide the means to live comfortably. Wealth reflects intergenerational transitions as well as accumulation of income and savings. Income, age, material status, family size, religion, occupation, and education are all predictors for wealth attainment.

CHAPTRE 4
ANALYSIS OF FACTORS AFFECTING ON POVERTY OF HOUSEHOLDS

In this section, the data is collected from the survey through questionnaire administration are organized, analyzed and interpreted in accordance with the methods and procedures.

4.1 Demographic and Socio-economic Characteristics of Households in Kyaukpadaung Township

The following tables are demographic and socio-economic characteristics of Households in Kyaukpadaung Township.

4.1.1 Gender of Household Heads

The following table shows gender of household heads in Kyaukpadaung Township.

Table 4.1 Gender of Household Heads

Gender	Frequency	Percent
Male	131	43.7
Female	169	56.3
Total	300	100

Source: Survey Data (2021)

According to the above table, there are 131 males household heads and 169 females household heads in 300 respondents which is 43.7 percent and 56.3 percent respectively. So, the numbers of female are more than the number of male.

4.1.2 Age of Household Heads

The following table is the age of household heads in Kyaukpadaung Township. According to table 4.1, the eldest head is 91 years old and the youngest head is 22 years old. Most of the household heads are between 22 and 31 years and there are 87 persons which is 29.0 percent. So, we can estimate that the most of household heads are contributed in labor force. Least of household heads are between 82 and 91 years and there are only one person which is 0.3 percent.

Table 4.2 Age of Household Heads

Age	Frequency	Percent
22-31	87	29.0
32-41	69	23.0
42-51	78	26.0
52-61	38	12.7
62-71	21	7.0
72-81	6	2.0
82-91	1	0.3
Total	300	100

Source: Survey Data (2021)

4.1.3 Education of Household Heads

The following table describes the education of Household heads in Kyaukpadaung Township.

Table 4.3 Education of Household Heads

Education	Frequency	Percent
Uneducated	9	3.0
Monastic education	1	0.3
Primary	62	20.7
Middle	96	32.0
High	77	25.7
University	9	3.0
Graduate	46	15.3
Total	300	100

Source: Survey Data (2021)

According to table 4.3, 32.0 percent of the household heads are middle education level. Monastic education is 0.3 percent. The uneducated household heads and university education level of the household heads are 3.0 percent respectively.

4.1.4 Occupation of Household Heads

The following table shows occupation of the household heads in Kyaukpadaung Township. According to table 4.4, 60.7 percent of household heads are random workers, 15.3 percent of household heads are dependent and 9.7 percent of household head's occupation are private owner & employee. There are 8.7 percent and 5.7 percent respectively of household head's occupation that are livestock worker and government employee.

Table 4.4 Occupation of Household Heads

Occupation	Frequency	Percent
Agriculture and livestock	26	8.7
Government employee	17	5.7
Private owner & employee	29	9.7
Random Worker	182	60.7
Dependent	46	15.3
Total	300	100

Source: Survey Data (2021)

4.1.5 Family Size of Households

The following table 4.1.5 shows the family size of households.

Table 4.5 Family Size of Households

Family Size	Frequency	Percent
1	12	4.0
2	74	24.7
3	75	25.0
4	67	22.3
5	40	13.3
6	22	7.3
7	8	2.7
8	1	0.3
9	1	0.3
Total	300	100

Source: Survey Data (2021)

According to table 4.5, 75 households have three family members and that percentage is 25 percent. Sixty-seven households have four family members and that percentage is 22.3 percent. There are two household that have eight and nine family members and the percentage is 0.3 percent respectively.

4.1.6 Number of Students of Households

The following table shows the number of students of Households.

Table 4.6 Number of Students of Households

Number of Students	Frequency	Percent
0	184	61.3
1	64	21.3
2	43	14.3
3	8	2.7
4	1	0.3
Total	300	100

Source: Survey Data (2021)

According to table 4.6, 184 households do not have students which is 61.3 percent. One household has four students which is 0.3 percent. It shows that there is only one household that has four students. And 64 households have only one student which is 21.3 percent.

4.2 Properties of Households in Kyaukpadaung Township

The following sections are shown by transportation facilities, entertainment facilities, loan condition, home ownership, housing type, sanitation, distance from important centers and source of drinking, source of energy, source of fuel and garbage system.

4.2.1 Ownership of Transportation Facilities in Kyaukpadaung Township

The table 4.7 shows ownership of transportation facilities in Kyaukpadaung Township. The table describes that most of the households have motorcycles and the percentage is 69.0 percent. Least of the households' are horsebox which is 1.3 percent. The owners of mini Oway and tricycle are 2.3 percent respectively.

Table 4.7 Transportation Facilities of Households

Categories	Yes		No	
	Households	Percent	Households	Percent
Car	19	6.3	281	93.7
Motorcycle	207	69.0	92	30.7
Bicycle	57	19.0	243	81.0
Mini Oway	7	2.3	293	97.7
Tricycle	7	2.3	293	97.7
Tricar	18	6.0	282	94.0
Horsebox	4	1.3	296	98.7

Source: Survey Data (2021)

4.2.2 Ownership of Entertainment Facilities in Kyaukpadaung Township

The following table shows ownership of entertainment facilities in Kyaukpadaung Township.

Table 4.8 Entertainment Facilities of Households

Categories	Yes		No	
	Households	Percent	Households	Percent
TV/ VCD	208	69.0	92	30.7
Radio	93	31.0	207	69.0

Source: Survey Data (2021*)*

According to the above table, 69.0 percent of households own TV and 31.0 percent of households own radio.

4.2.3 Ownership of Home Appliances of the Households in Kyaukpadaung Township

The table 4.9 shows the ownership of home appliances of the households in Kyaukpadaung Township. According to the table 4.9, 98.3 percent of households own telephone, 70.3 percent of households own rice cooker, 37 percent of households own refrigerator and 3.3 percent of households own gas stove.

Table 4.9 Ownership of Home Appliances of the Households

Categories	Yes		No	
	Households	Percent	Households	Percent
Sewing Machine	31	10.3	269	89.7
Refrigerator	111	37.0	189	63.0
Rice Cooker/ Iron	211	70.3	89	29.7
Mobile Phone	296	98.3	4	1.3
Gas Stove	10	3.3	290	96.7
Other	21	3.8	279	96.2

Source: Survey Data (2021)

4.2.4 Home Ownership of Households in Kyaukpadaung Township

The following table shows home ownership of Households in Kyaukpadaung Township. According to the table, almost all of the households have own house (297) and the other 3 households are renter.

Table 4.10 Home Ownership of Households

Home Ownership	Number of households	Percent
Own	297	99
Rent	3	1
Total	300	100

Source: Survey Data (2021)

4.2.5 Housing Type of Households in Kyaukpadaung Township

Table 4.11 shows the housing types of households.

Table 4.11 Housing Type of Households

Housing Type	Number of households	percent
RC	20	6.7
Brick	80	26.7
Wood	100	33.3
Bamboo	100	33.3
Total	300	100

Source: Survey Data (2021)

According to the above table, most of the houses are wood and bamboo houses and it is 33.3 percent respectively. The brick houses are 26.7 percent and the other 6.7 percent are RC houses.

4.2.6 Types of Sanitation Condition of Households in Kyaukpadaung Township

This table shows the sanitation condition of households.

Table 4.12 Types of Sanitation Used by Households

Types of Sanitation	Number of households	Percent
Using water toilet	108	36.0
Tradition	189	63.0
No	3	1.0
Total	300	100.0

Source: Survey Data (2021)

According to table 4.12, most of the households use the traditional toilets, it is 63.0 percent. Another 36.0 percent of households use water toilets and only 1 percent has no toilet.

4.2.7 Accessibility to Important Centers of Households in Kyaukpadaung Township

The following table shows accessibility to important centers of households in Kyaukpadaung Township.

Table 4.13 Accessibility to Important Centers of Households

Type	Far	Percent	Near	Percent
School	97	32.3	203	67.7
Bazaar	168	56.0	132	44.0
Clinic	142	47.3	158	52.7
Hospital	253	84.3	46	15.3

Source: Survey Data (2021)

According to the above table, 203 households answered “near from school”. One hundred and fifty eight households answered “near from clinic”. All households answered “far away from bazaar and far away from hospital”.

4.2.8 Source of Drinking Water of Households in Kyaukpadaung Township

The following table shows source of drinking water of households.

Table 4.14 Source of Drinking Water of Households

Source of Drinking Water	Number of households	Percent
Fresh Water	203	67.7
Well	12	4.0
Tube Well	33	11.0
River	20	6.7
Lake	16	5.3
Spring Water	16	5.3
Total	300	100

Source: Survey Data (2021)

According to the above table, most of the households got from purified water, it is 67.7 percent. Only twelve households got from tube well. The households got from lake and spring water are 5.3 percent respectively.

4.2.9 Source of Energy of Households in Kyaukpadaung Township

The following table shows source of energy of households.

Table 4.15 Source of Energy of Households

Types of Energy	Number of households	Percent
Electricity	246	74.3
Battery	51	24.3
Solar	2	1.3
Total	300	100.0

Source: Survey Data (2021)

According to the above table, 74.3 percent of households use electricity, 24.3 percent of households use battery and only 1.3 percent of household use battery. It shows that the source of energy condition of this township is good and at high level.

4.2.10 Source of Fuel for Cooking of Households in Kyaukpadaung Township

The following table shows source of fuel for cooking of households.

Table 4.16 Source of Fuel for Cooking of Households

Types of Cooking	Number of households	Percent
Electric	223	74.3
Fire Wood /Charcoal	73	24.3
Gas	4	1.3
Total	300	100.0

Source: Survey Data (2021)

According to the above table, 74.3 percent of households use electric, 24.3 percent of households use fire wood and only 1.3 percent use gas. Therefore, it can be assumed that the families of this township use fire wood for cooking.

4.2.11 Garbage System of Households in Kyaukpadaung Township

Table 4.17 shows the garbage system of households.

Table 4.17 Garbage System of Households

Garbage System	Number of household	Percent
Waste bin/ Rubbish car/Cart	159	53.0
Garbage Fire	96	32.0
River or stream	23	7.7
No space	20	6.7
Other	2	0.7
Total	300	100.0

Source: Survey Data (2021)

According to the above table, 159 of households do rubbish car which percentage is 53 percent , 96 of households do garbage fire which percentage is 32 percent and 23 households do river or stream which percentage is 7.7 percent.

4.3 Economic Conditions of Households in Kyaukpadaung Township

Economic condition of Households in Kyaukpadaung Township is shown with loan conditions of households, monthly income of household and monthly expenditure according to the survey data of this township.

4.3.1 Loan Conditions of Households in Kyaukpadaung Township

The following table shows the loan conditions of households.

Table 4.18 Loans of Households

Loan conditions of Households	Number of Households	Percent
Yes	76	25.3
No	224	74.7
Total	300	100

Source: Survey Data (2021)

According to the above table, the household who has loan are 76 households and it is 25.3 percent. Remaining 224 households do not have loan and it is 74.7 percent.

4.3.2 Monthly Income of Households

The following table shows the monthly income for households. According to the table, the monthly income of 179 households is under 375000 Kyats. At least monthly income is Kyats 90000. Only 2 households are between 1515000 Kyats and 1799999 Kyats and 3 households are between 1230000 Kyats and 1514999 Kyats.

Table 4.19 Monthly Income for Households

Income (Kyats)	Number of households	Percent
90000-374999	179	59.7
375000-659999	94	31.3
660000-944999	16	5.3
945000-1229999	6	2.0
1230000-1514999	3	1.0
1515000-1799999	2	0.7
Total	300	100.0

Source: Survey Data (2021)

4.3.2 Monthly Expenditure of Households

This table shows monthly expenditure for households.

Table 4.20 Monthly Expenditure for Households

Expenditure (Kyats)	Number of households	Percent
22100- 162099	92	30.7
162100- 302099	150	50.0
302100- 442099	40	13.3
442100- 582099	12	4.0
582100- 722099	3	1.0
722100-862099	3	1.0
Total	300	100.0

Source: Survey Data (2021)

According to the table 4.20, the monthly expenditure of 150 households is under 302100 Kyats. At least monthly expenditure is Kyats 22100. The monthly expenditure of 3 households is between 722100 and 862099 Kyats.

4.4 Effects of Factors on Expenditures of Households in Kyaukpadaung Township

The multiple regression analysis was applied to investigate the factors of monthly expenditure in sample households of Kyaukpadaung Township. To develop the multiple regression model, the monthly expenditure of households was used as dependent variable and household heads age, total income, number of students in each household were used as independent variables.

The estimated multiple regression model is

$$\widehat{Y}_i = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \quad (4.1)$$

Where,

Y_i = Monthly expenditure of households

X_1 = Total income

X_2 = Number of students

X_3 = Household heads age

Table 4.21 Results of Multiple Regression Model

Independent variable	Coefficient	Standard error	t	Sig	VIF
Constant	53556.779	12736.960	4.205	0.000	
Total income	0.357***	0.019	18.866	0.000	1.049
Number of Students	28105.313***	5506.223	5.104	0.000	1.099
Household heads age	9731.139***	3414.203	2.850	0.005	1.050
Adjusted R^2	0.595				
F-value	147.639				

Source: Survey Data (2021)

Dependent variable: Monthly expenditure

*** denotes significant at 1 percent level, ** denotes significant at 5 percent level,

* denotes significant at 10 percent level.

Multiple Regression equation is

$$\hat{Y} = 53556.779 + 0.711X_1 + 0.197X_2 + 0.107X_3 \quad (4.2)$$

Results show that F value is 147.639 that is significant at p=0.000(<0.01),

suggesting that independent variables have significantly. Adjusted R^2 is 0.595. It had been found that total income, number of students and household heads age are statistically significance at 1 percent level respectively. The result shows that average monthly expenditure of households is 53556.779 MMK. The multiple regression equation shows that, monthly expenditure is expected to increase by 0.357 MMK, if total income increases by 1 Kyats. If number of students increased by 1 member, monthly expenditure is increased by 28105.313 MMK. If household head age is increased by 1 year, monthly expenditure is increased by 9731.139 MMK. The regression coefficient between total income and total expenditure is 0.357 (t=18.866, p=0.000).This shows that there is direct relationship between total income and monthly expenditure. The regression coefficient between number of students and monthly expenditure is 28105.313 (t=5.104, p=0.000).This shows that there is direct relationship between number of family members and monthly expenditure. The regression coefficient between household head age and monthly expenditure is 9731.139 (t = 2.850, p = 0.005).This shows that there is direct relationship between household head age and total expenditure.

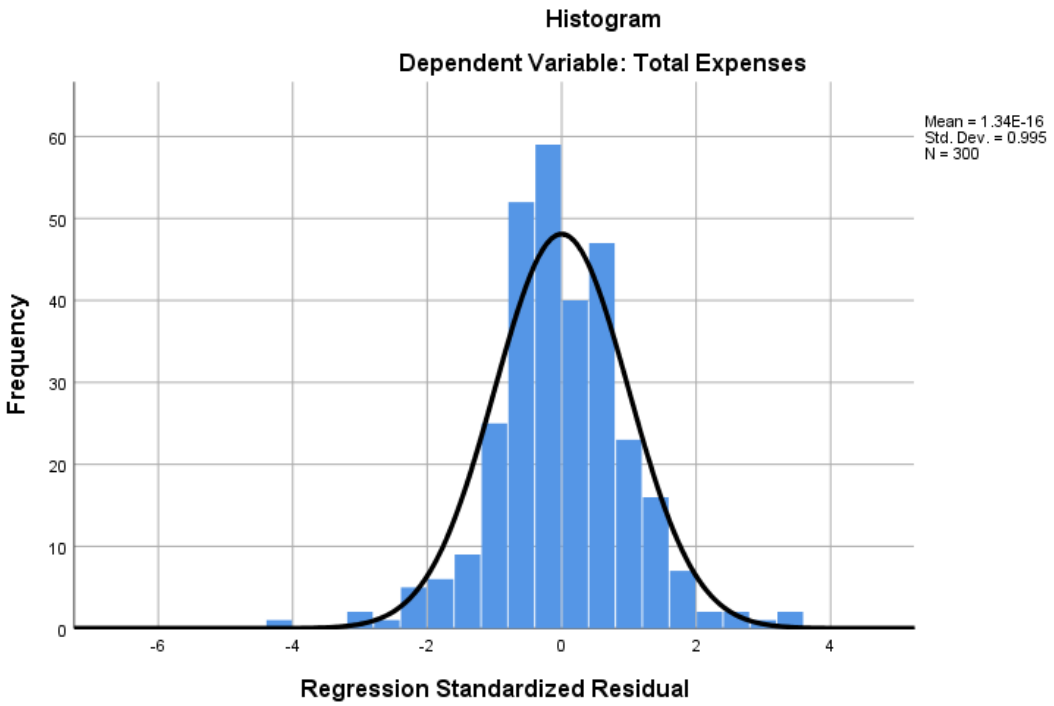
4.5 Testing for the Assumptions about Multiple Regression

To determine the violation of required assumption from multiple linear regression model for socio-economic status of households in Kyaukpadaung Township, the following procedures have been used.

(i) Testing for Normality of Disturbances

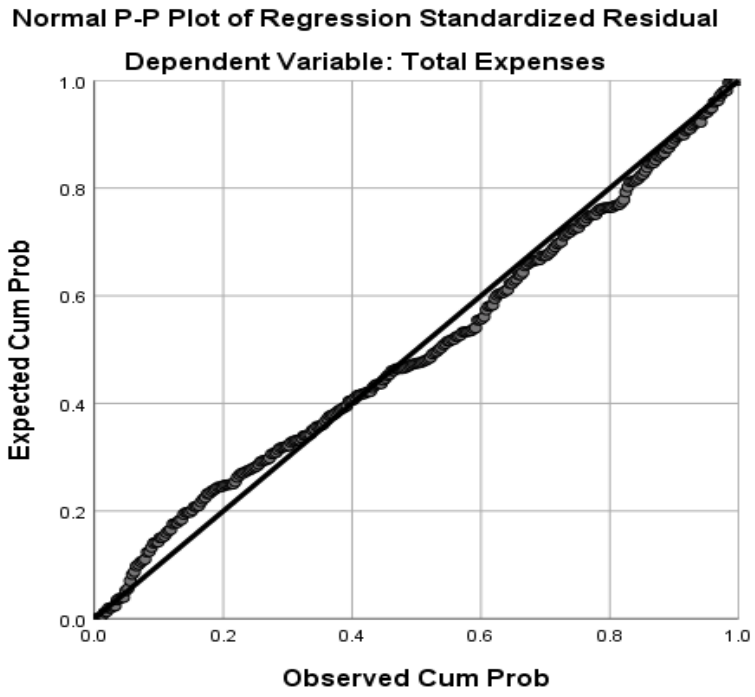
One of the basic assumptions is that disturbances are normally distributed with zero mean and constant variance. To check whether the disturbances are normally distributed, histogram and Normal P-P plot of the disturbances can be constructed. The histogram of the residual and Normal P-P plot for socio-economic position of households in Kyaukpadaung Township are shown in Figure 4.1 and 4.2.

The histogram in Figure 4.1 appears to be pile fashioned. Similarly, the normal probability plots is virtually a straight line. Although the graphs do not provide formal statistical test of normality, it will provide a descriptive display. According to histogram and Normal P-P plot, it can be concluded that the normality assumption appears to be generally reasonable.



Source: Survey Data (2021)

Figure 4.1 Histogram for Residual

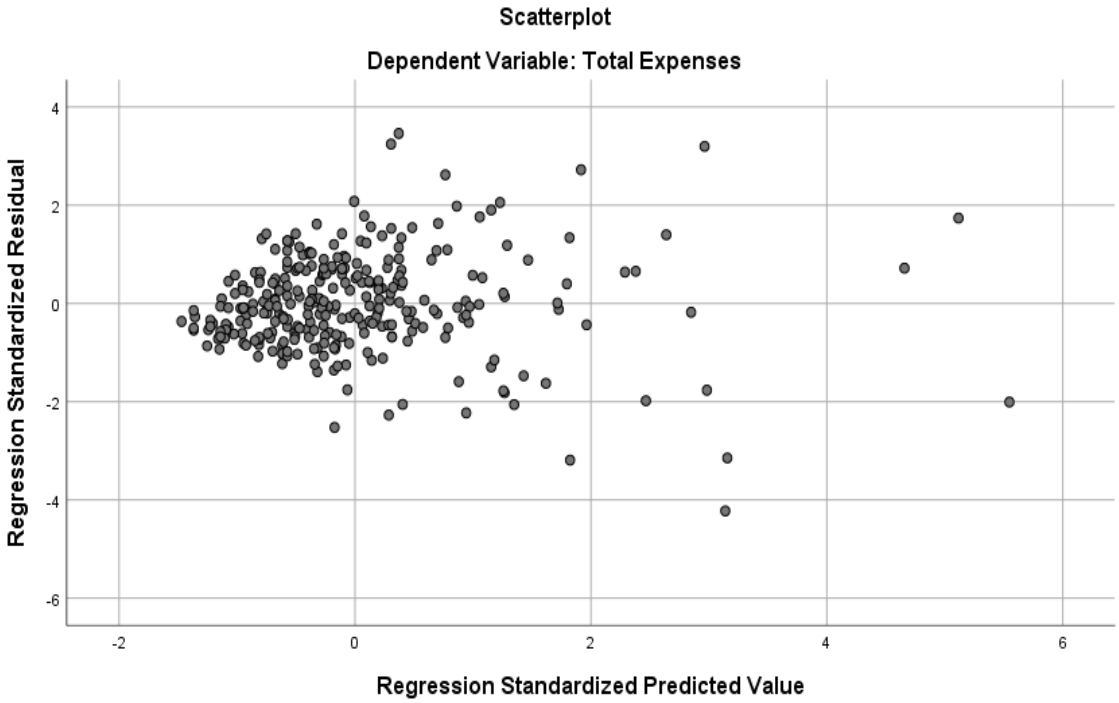


Source: Survey Data (2021)

Figure 4.2 Normal P-P Plot of Regression Residuals

(ii) Testing for the Presence of Heteroscedasticity Problem

The two figures are error plots. The White test is to be used in this study to detect the presence of heteroscedasticity. Another basic assumption of the multiple regression. Regression model is homoscedasticity. In the presence of heteroscedasticity, the regression coefficients become less efficient. Heteroscedasticity can often be detected by plotting the estimated Y values against the disturbances. If any pattern is displayed, heteroscedasticity is likely present. Figure 4.3 represents the predicted on expenditure of household on X-axis and the residual values on Y axis.



Source: Survey Data (2021)

Figure 4.3 Residuals Pattern for Heteroscedasticity Expenditure of Households

This figure shows that heteroscedasticity appears to be absent.

(iii) Detecting Multicollinearity

Multicollinearity arises when one of the independent variables is linearly related to one or more of the other independent variables. Such a situation violates one of the assumptions for multiple regression. Specifically, multicollinearity occurs if there is a high correlation between two independent variables.

To detect multicollinearity is to use the variance inflation factor (VIF). It is measured the degree of multicollinearity contributed by independent variable. In the

multiple regression model, the VIF for total income, number of students and household head age are 1.049, 1.099 and 1.050 respectively. The sum of VIF for these independent variables is 3.198. Since the sum of the VIF is less than 10, then it is concluded that multicollinearity is not serious problem in the multiple regression model for socio-economic status of household.

4.6 Factors Affecting on Wealth Index of Socio-economic Status of Households in Kyaukpadaung Township

The following table shows interpretation of results from principal component analysis (PCA). The KMO measure of sampling adequacy tests whether the partial correlation among items is small. The Kaiser-Meyer-Olkin measure of sampling adequacy varies between 0 and 1, and the values closer to 1 are better. A value of greater than 0.5 is suggested to be the minimum (Field, 2005). The KMO in table 4.22 is 0.777 which is above 0.5 is satisfactory. Bartlett's Test of Sphericity helps test the null hypothesis that the correlation matrix is an identity matrix. An identity matrix is a matrix in which all the diagonal elements are 1 and all off diagonal elements are 0. According to table 4.22, the Bartlett's Test of Sphericity was significant at 0.000, which means there was a relationship between the variables includes in the analysis. The significant level was small enough to reject the null hypothesis, which means that the correlation matrix was not an identity matrix. The Bartlett's Sphericity test and the KMO index to detect if the researcher can or cannot summarize the information provided by the initial variables in a few number of factors.

Table 4.22 KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.777
Bartlett's Test of Sphericity	Approx. Chi-Square	417.182
	Df	36
	Sig.	0.000

Source: Survey Data (2021)

4.6.1 Total Variance Explained by Factors

The sample variation of the components has been described in table 4.23. The following table reports the variance explained by each component as well as the

cumulative variance explained by all the components. Table 4.23 shows the amount of variance in the total correlation of variables which is explained by the components. Component 1 explain 30.174 percent of the variance in the item, component 2 to component 3 explain 12.860 percent and 12.387 percent of the variance in the items in the component respectively. The cumulative percentage column contains the cumulative percentage of the variance accounted for by the current and all preceding components. According to table 4.23 the 3rd row shows a value of 12.387 percent of the total variance.

Table 4.23 Total Variance Explained by Factors

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	Percent of Variance	Cumulative Percent	Total	Percent of Variance	Cumulative Percent
1	2.8000	31.113	31.113	2.716	30.174	30.174
2	1.176	13.062	44.176	1.157	12.860	43.034
3	1.012	11.245	55.421	1.115	12.387	55.421
4	0.926	10.289	65.710			

Source: Survey Data (2021)

4.6.3 Wealth Index of Households in Kyaukpadaung Township

The following table shows Wealth Index of Households in Kyaukpadaung Township. According to the table, almost households of 21.3 percent are poorest and 19.7 percent of households are rich. The others poorer and middle are 20.0 percent respectively.

Table 4.24 Wealth Index of Households

Wealth Index	Frequency	Percent
Poorest	64	21.3
Poorer	60	20.0
Middle	60	20.0
Rich	59	19.7
Richest	57	19.0
Total	300	100.0

*Source***:** Survey Data (2021)

4.6.5 Model Fitting Information of Households in Kyaukpadaung Township

The results of the overall model evaluation binary logistic regression model are presented in table 4.25.

Table 4.25 Model Fitting Information of Households

Model Fitting Criteria	χ^2 value	df	*p*-value
Omnibus Tests of Model Coefficient	78.854	7	0.000
Hosmer and Lemeshow (H-L) Tests	6.897	8	0.548
Cox & Snell R^2	0.231		
Nagelkerke R^2	0.365		
Overall percentage	79.7 percent		

Source: Survey Data (2021)

According to the Omnibus tests of model coefficients gives a Chi-Square of 78.854 on 7 df significant beyond 0.000. There is no evidence of lack of fit based on the H-L statistic (Chi-Square = 6.897, df = 8, p = 0.548). The model fitting information includes two different ways of estimating R square (Cox & Snell R^2 and Nagelkerke R^2). Overall 79.7 percent of the poverty of households is predicted correctly.

4.6.6 Summary Result for the Binary logistic Regression Model of Households in Kyaukpadaung Township

Factors affecting on poverty of households are analyzed by using binary logistic regression model. In this model, the total expenditures of household are dependent variable. The binary logistic regression model of socio-economic status can be described as follows.

$$\text{Logit}\,(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon \qquad (4.3)$$

Poverty of households in Kyaukpadaung township usage is dependent variable and it was given 1 if the poor of the household and 0 if non-poor condition of household.

Y = Dependent variables

In this study, Y = 1, if household is poor

Y = 0, if household is non-poor

Gender, age, occupation, number of family members, number of students and total income are used as independent variables. These variables are categorized as followed:

ΔX = Vector of independent variables

X_1 = Age

X_2= Gender = 1 if household heads is male

= 0 if household heads is female

X_3= Occupation = 1 if household heads occupation is agriculture

= 0 if household heads occupation is other

X_4 = Number of family members

X_5 = Number of students

X_6 = Total income

The summary result for the binary logistic regression model of households in Kyaukpadaung Township is showed in table 4.26.

Table 4.26 Summary Result for the Binary Logistic Regression Model of Households

	B	S.E	Wald	df	Sig	Exp(B)
Constant	1.330	0.902	2.176	1	0.040	3.783
Age	0.033*	0.134	0.060	1	0.077	1.033
Gender (1)	0.368***	0.349	1.116	1	0.009	1.445
Occupation(1)	-0.044***	0.542	0.007	1	0.000	0.957
Family size	-0.188	0.168	1.249	1	0.264	1.207
No. of Student	-0.070*	0.253	0.076	1	0.083	0.933
Total income	0.000***	0.000	38.194	1	0.000	1.000
Loans	-0.346***	0.431	0.645	1	0.022	0.708

Source: Survey data (2021)

*** denotes significant at 1 percent level, ** denotes significant at 5 percent level,

* denotes significant at 10 percent level`

It has been found that the coefficient of gender, occupation, total income and loans are statistically significant at 1 percent level. The coefficient of age and number of students are statistically significant at 10 percent level. Age, gender and total income are positive sign. Number of students, occupation and loans are negative sign. If the male household heads are more increasing, the poorer and poorest level of household is more increasing 0.368. If there is more increasing in the age of household heads as 1 year, the poorer and poorest level of households is more increasing in 0.033. If there is more increasing in the family size as 1 person, the poorer and poorest level of household is more decreasing in 0.188. If there is more increasing in the student size as 1 person,

the poorer and poorest level of households is more decreasing in 0.070. But if there is more increasing in the borrowing as 1 MMK, the poorer and poorest level of households is more decreasing in 0.346 MMK. If there is more increasing in the occupation of household's head which is agriculture, the poorer and poorest level of household heads is more decreasing in 0.044.

CHAPTER 5
CONCLUSION

This chapter focuses on the conclusion of the thesis related to findings, suggestions and recommendations and needs for further study.

5.1 Findings

In this study of factors affecting on poverty of household in Kyaukpadaung Township according to the objective one, it can be found that there are more females than males. The number of persons in working age group was larger than that of dependents in this township. One can also say that the volume of labour force was high in this township.

In the educational attainment of Kyaukpadaung Township, the most population are middle education level. The most households in Kyaukpadaung Township owns wood and bamboo house and live by their own houses. Most of households used the traditional toilet. Therefore, it can be seen that the sanitation condition was fairly good condition in Kyaukpadaung Township.

The most households had available electricity for main source of lighting and the least households use battery for source of lighting in Kyaukpadaung Township. Most of the households got water for drinking from fresh purified water. Most of the households used electric for cooking. Therefore, it can be assumed that the types of cooking condition was good condition in Kyaukpadaung Township. In the study of the households by availability and related amenities, it is found that the most available item for amenities is television. Most of the households used motorcycle. Thus, the socio-economic positions of households in Kyaukpadaung Township have fairly good condition.

The household income level in the study area, it is found that most of the sample households have monthly income of Kyats 90000 - 374999. Thus, the economic condition of the sample households in the study area is fairly good. It is found that 59.7 percent of the sample households in the study area.

According to multiple regression model, the relationship between the total expenses and the total income of households are direct relation. If there is more increasing in the total income of households as 1 kyat, the total expenses of households is more increasing in 0.357 kyat. The expenditure of households are depend on income

of households, age of household heads, education of household heads, occupation, family size and student size.

Binary logistic regression model is applied to find the relationship between wealth index and the independent variables. When the poverty level is considered as dependent variable, the poverty level can be found more than five categories such as poorest, poorer, middle, rich and richest. According to the binary logistic regression model, the coefficient of gender, occupation, total income and loans are statistically significant at 1 percent level. Total income is not practically significant, although statistically significant. The coefficient of age and number of students are statistically significant. Age, gender and total income are positive sign. Therefore, increasing the male household heads and age of household heads cause to increase the poorest situation of households.

Number of students, occupation and loans are negative sign. If there is more increasing in the student size, the occupation of household's head which is agriculture and borrowing of the loan, the poorest situation of household is more decreasing. Therefore, getting occupation, increasing the number of students and loans are effect to decrease the poorest situation on households.

5.2 Suggestions and Recommendations

Almost all of the people in the township are needed to improve the living standard. The involvement of people in social organizations plays an important role in improvement of social environment and also social welfare. Though the involvement of social organizations in this township is less. Thus, the people must be encouraged to involve in social organizations and to emerge more social organizations in Kyaukpadaung Township.

To achieve all round good socio-economic position of households in Kyaukpadaung Township, there should has special development schemes and implement them efficiently for the overall benefits of Kyaukpadaung Township. As a few households in Kyaukpadaung Township discard garbage fire, it can be air pollution. Education opportunities need to be more supported in there. Moreover, health care is not safe for households because of all four wards are far from hospital. As there are households that still use battery for lighting, electric should be fulfilled for households that still use battery.

To resolve the poverty problem, the households in the four wards should be reduced the gap between incomes and expenditures, whether by increasing incomes or reducing the cost of living, or by a proper mix of the two. To increase income, the government should prepare new job opportunities (full times and part times) and should equal pay for men and women. The government should arrange short course for profession that support small own business by using loans. And this may help also to increase income. For development of socio-economic status, transportation should be better to facilitate. If the more economic development in the region, poverty can be declined and urban and rural economic will develop.

5.3 Needs for Further Study

This study is focused on factors affecting on poverty of households in Kyaukpadaung Township by using face to face interview for data collection. In this study, poverty of households in Kyaukpadaung Township are determined based on wealth index. If food poverty line is used, the results will be more comprehensive. If the study is focused on another township instead of Kyaukpadaung Township, the result will be more interested. From the socio-demographic factors, occupation, number of students, family size and gender of household heads are only considered. If other important variable are studied to determine poverty, the results will be more interested.

REFERENCES

Ahmad, A. A. (2015). Economic Wealth Index: A tool to study the Economic Health of Districts of selected States in India. *International Journal of Scientific and Research Publications*

Anyanwu, J. C. (2013). *Determining the correlates of poverty for inclusive growth in Africa*. Tunisia: African Development Bank

Azpitarte, F. (2010). Measuring poverty using both income and wealth: An empirical comparison of multidimensional approaches using data for the US and Spain

Beegle, K., Christiaensen, L., Dabalen, A., & Gaddis, I. (2016). *Poverty in a rising Africa*. The World Bank.

Besley, T., & Burgess, R. (2003). Halving global poverty. *Journal of economic perspectives*, *17*(3), 3-22.

Block, S. A., Masters, W. A., & Bhagowalia, P. (2010). *Child Undernutrition, Household Poverty and National Income in Developing Countries: Quantile Regression Results* (No. 320-2016-10039)

Cahyat, A., Gönner, C., & Haug, M. (2007). *Assessing household poverty and wellbeing: a manual with examples from Kutai Barat, Indonesia*. CIFOR.

Cardoso, E. (1992). *Inflation and poverty* (No. w4006). National Bureau of Economic Research

Carter, M. R., & Barrett, C. B. (2006). The economics of poverty traps and persistent poverty: An asset-based approach. *The Journal of Development Studies*, *42*(2), 178-199.

Chani, D., Irfan, M., Pervaiz, Z., Jan, S. A., Ali, A., & Chaudhary, A. R. (2011). Poverty, inflation and economic growth: empirical evidence from Pakistan. *World Applied Sciences Journal*, *14*(7), 1058-1063

Chasekwa, B., Maluccio, J. A., Ntozini, R., Moulton, L. H., Wu, F., Smith, L. E, ... & SHINE Trial Team. (2018). Measuring wealth in rural communities: lessons from the Sanitation, Hygiene, Infant Nutrition Efficacy (SHINE) trial. *PLoS one*, *13*(6), e0199393.

Dahiya, B. (2012). Cities in Asia, 2012: Demographics, economics, poverty, environment and governance. *Cities*, *29*, S44-S61

Dekker, M. (2006). Estimating wealth effects without expenditure data: Evidence from rural Ethiopia. *Ethiopian Journal of Economics*, *15*(1), 35-53.

Douidich, M., Ezzrari, A., Van der Weide, R., & Verme, P. (2016). Estimating quarterly poverty rates using labor force surveys: a primer. *The World Bank Economic Review*, *30*(3), 475-500

Ergo, A., Htoo, T. S., Badiani-Magnusson, R., & Royono, R. (2019). A new hope: from Neglect of the health sector to aspirations for Universal Health Coverage in Myanmar. *Health policy and planning*, *34*(Supplement_1), i38-i46.

Gjoni, A. A. A. E., Bedjaoui, A. A. R. A., Mangueira, A. A. D. S. A., Graca, J., Sturzenegger, A. L. A. C. F., Hayrapetyan, A. V. G. A., & Azam, S. (2015). Governors and Alternates of the World Bank| June 30, 2017. *Development*, *11*.

Gostin, L. O., & Friedman, E. A. (2019). Imagining Global Health with Justice: Transformative Ideas for Health and Well-Being While Leaving No One Behind. *Geo. LJ*, *108*, 1535

Habyarimana, F., Zewotir, T., & Ramroop, S. (2015). Analysis of demographic and health survey to measure poverty of household in Rwanda. *African population studies*, *29*(1), 1472-1482

Hayati, D., Karami, E., & Slee, B. (2006). Combining qualitative and quantitative Methods in the measurement of rural poverty: the case of Iran. *Social Indicators Research*, *75*(3), 361-394

Jouini, N., Lustig, N., Moummi, A., & Shimeles, A. (2018). Fiscal Policy, Income Redistribution, and Poverty Reduction: Evidence from Tunisia. *Review of Income and Wealth*, *64*, S225-S248.

Klasen, S., & Misselhorn, M. (2008). *Determinants of the growth semi-elasticity of poverty reduction* (No. 176). Discussion papers, Ibero America Institute for Economic Research.

Lustig, N. (2017). Fiscal policy, income redistribution and poverty reduction in low and middle income countries. *Center for global development working paper*, (448).

Mahumud, R. A., Sultana, M., & Sarker, A. R. (2017). Distribution and determinants of low birth weight in developing countries. *Journal of preventive medicine and public health*, *50*(1), 18.

Marangio, R. (2012). The Somali crisis: Failed state and international interventions

McKernan, S. M., & Ratcliffe, C. E. (2002). Transition events in the dynamics of

poverty. *The Urban Institute, September*.

Mohanty, S. K. (2009). Alternative wealth indices and health estimates in India. *Genus*, *65*(2), 113-137

Nussbaumer, P., Bazilian, M., & Modi, V. (2012). Measuring energy poverty: Focusing on what matters. *Renewable and Sustainable Energy Reviews*, *16*(1), 231-243

Rutstein, S. O., & Staveteig, S. (2014). *Making the demographic and health surveys wealth index comparable* (Vol. 9). Rockville, MD: ICF International

Schmitt-Degenhardt, S. (2013). A regional perspective on poverty in Myanmar. *Yangon: UNDP*

Wanka, F. A. (2014). *The impact of educational attainment on household poverty in South Africa: A case study of Limpopo Province* (Doctoral dissertation, University of the Western Cape).

Warr, P. (2020). POVERTY AND INEQUALITY IN MYANMAR: 2005 TO 2017. *Living with Myanmar*, 125

Watson, D., Kehoe, S. H., Erzse, A., Compaoré, A., Debpuur, C., Nonterah, E. A. ... & Barker, M. (2020). Community perspectives of maternal and child health during nutrition and economic transition in sub-Saharan Africa. *Public Health Nutr*, 1-9

Win, S.L (2019). R for basic statistics and R for advanced statistics.

https://www.mm.undp.org/content/myanmar/en/home/countryinfo.html

https://www2.gnb.ca/content/gnb/en/departments/esic/overview/content/what_is_poverty.htm

APPENDIX (A)

QUESTIONNAIRE

FACTORS AFFECTING ON POVERTY OF HOUSEHOLDS IN KYAUKPADAUNG TOWNSHIP

Date ……..... Questionnaire No. ……

1. Address

Township…… Village……

2. Respondent

(a) Name…… (b) Age ……

(c) Male/Female……. (d) Education……..

(e) Occupation……. (f) Kinship with householder……

3. Head of Household

(a) Name……. (b) Age…… (c) Male/Female……

(d) Education…… (e) Occupation……

4. Family Topics

(a) Number of family

Sir No.	Name	Gender	Religion	Relation with Household Head	Age	Education	Marital Status	Occupation	Income

(b) Number of Students

Level of Education	No. of Students		Age	Type of school	Total
	Male	Female			
Primary					
Middle					
High					
University					
Total					

5. Properties of Sample Households

(a) Car (mini) ☐ (b) Mini Oway (Own) ☐ (c) Cycle ☐
(Truck) ☐ (Rent) ☐ ☐
(d) Bicycle ☐ (e) Tri cycle ☐ (f) Tri Car
(g) Horsebox ☐ (h) Refrigerator ☐ (i) TV, VCD, DVD, ☐
EVD ☐
(j) Telephone ☐ (k) Rice Cooker, Iron ☐ (l) Radio, Cassette
(m) Sewing machine ☐ (n) Gas stove ☐ (o) others ☐

6. Expenditures

No.	Type of Expenditure				Expenses (kyats)		
					One week	One month	One year
1.	Expenses for kitchen		Price	Amount			
	1.	Rice					
	2.	Oil					
	3.	Market (one week)					
2.	Fruits/Beverages						
3.	Education						
4.	Repair Cost						
5.	Recreation						
6.	Social cost						
7.	Health cost						
8.	General Cost						
Total							

7. Did you get loan? □ Yes □ No

(a) If you get loan; describe detail

Name of organization	Amount of loan	Interest	Reason of getting loan	Period

(b)Does income support by getting loan? □ Yes □ No

(c) If not support; please tick the following:

Less amount	High interest rate	Short loan period	Not use with correctly

(d) If you get support; describe detail

Name of organization	Reasons for support	Amount	Number of times	Period

8. Housing Condition

(a) Own □ (b) Rent □ (c) Others □

Housing Type	R.C	Brick	Wooden	Bamboo
Toilet Type		Cover Pit Toilet	Open pit Toilet	No

9. Distance Condition

Distance	Near	Far
School from home		
Shop from home		
Clinic from home		
Hospital from home		

10. Drinking Water

Please tick	Purified drinking water	Well	Tube well	River	Lake	Others

11. Condition of Cooking

Please tick	Electricity	Wooden	Charcoal	Gas	Others

12. Energy Condition

Please tick	Electricity	Battery	Solar	Generator	Others

13. Garbage System

Please tick	Garbage cans, garbage trucks, garbage carts	Fire/Underground	River	No Stable	Others

14. What do you want to do to improve your social life?

……

……

APPENDIX (B)

REGRESSION

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.774[a]	0.599	0.595	77610.932

a. Predictors: (Constant), Age, Total Income, No. of Student
b. Dependent Variable: Total Expenses

ANOVA[a]

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	2667895469081.90	3	889298489693.968	147.639	0.000[b]
Residual	1782943194356.76	296	6023456737.692		
Total	4450838663438.66	299			

a. Dependent Variable: Total Expenses
b. Predictors: (Constant), Age, Total Income, No. of Student

Coefficients[a]

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
1 (Constant)	53556.779	12736.960		4.205	0.000		
Total income	0.357	0.019	0.711	18.866	0.000	0.953	1.049
No. of Student	28105.313	5506.223	0.197	5.104	0.000	0.910	1.099
Age	9731.139	3414.203	0.107	2.850	0.000	0.952	1.050

c. Dependent Variable: Total Expenses

FACTOR ANALYSIS

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.777
Bartlett's Test of Sphericity	Approx. Chi-Square	417.182
	df	36
	Sig	0.000

Total Variance Explained

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.800	31.113	31.113	2.716	30.174	30.174
2	1.176	13.062	44.176	1.157	12.860	43.034
3	1.012	11.245	55.421	1.115	12.387	55.421
4	0.926	10.289	65.710			
5	0.792	8.795	74.505			
6	0.711	7.900	82.405			
7	0.655	7.278	89.683			
8	0.504	5.603	95.286			
9	0.424	4.714	100.000			

LOGISTIC REGRESSION

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	78.854	7	0.000
	Block	78.854	7	0.000
	Model	78.854	7	0.00

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	221.388[a]	0.231	0.365

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6.897	8	0.548

Percentile Group of Wealth

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Poorest	64	21.3	21.3	21.3
	Poor	60	20.0	20.0	40.3
	Middle	60	20.0	20.0	60.3
	Rich	59	19.7	19.7	80.3
	Richest	57	19.0	19.0	100.0
	Total	300	100.0	100.0	

Variable in the Equation

	B	S.E	Wald	df	Sig	Exp(B)
Constant	1.330	0.902	2.176	1	0.040	3.783
Age	0.033*	0.134	0.060	1	0.077	1.033
Gender(1)	0.368***	0.349	1.116	1	0.009	1.445
occupation(1)	-0.044***	0.542	0.007	1	0.000	0.957
Family size	-0.188	0.168	1.249	1	0.264	1.207
No. of Student	-0.070*	0.253	0.076	1	0.083	0.933
Total income	0.000***	0.000	38.194	1	0.000	1.000
Loans	-0.346***	0.431	0.645	1	0.022	0.708