

**YANGON UNIVERSITY OF ECONOMICS
DEPARTMENT OF STATISTICS**

**PRINCIPAL COMPONENT ANALYSIS
OF INFANT MORTALITY IN MYANMAR**

HSU HLAING PHYO

M.Econ (Statistics)

Roll No.1

MAY, 2018

**YANGON UNIVERSITY OF ECONOMICS
DEPARTMENT OF STATISTICS**

**PRINCIPAL COMPONENT ANALYSIS
OF INFANT MORTALITY IN MYANMAR**

This thesis is submitted as a partial fulfillment towards
the Degree of Master of Economics (Statistics)

HSU HLAING PHYO
M.Econ (Statistics)
Roll No.1

MAY, 2018

**YANGON UNIVERSITY OF ECONOMICS
DEPARTMENT OF STATISTICS**

**PRINCIPAL COMPONENT ANALYSIS
OF INFANT MORTALITY IN MYANMAR**

Thesis submitted as a partial fulfillment towards the
Degree of Master of Economics (Statistics)

**BY
HSU HLAING PHYO
M.Econ (Statistics)
Roll No.1**

MAY, 2018

**YANGON UNIVERSITY OF ECONOMICS
DEPARTMENT OF STATISTICS**

**PRINCIPAL COMPONENT ANALYSIS
OF INFANT MORTALITY IN MYANMAR**

This thesis is submitted to Board of Examination as partial fulfillment of the
requirement for degree of Master of Economics (Statistics)

Approved by the Board of Examiners

Submitted by:



Hsu Hlaing Phyio
Roll No.1
M.Econ (Statistics)

Supervised by:



Daw Aye Aye Maw
Lecturer
Department of Statistics
Yangon University of Economics

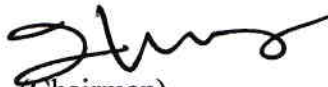
MAY, 2018

YANGON UNIVERSITY OF ECONOMICS

DEPARTMENT OF STATISTICS

This is to certify that thesis entitled **“PRINCIPAL COMPONENT ANALYSIS OF INFANT MORTALITY IN MYANMAR”**, submitted as a partial fulfillment towards the requirements of the degree of Master of Economics (Statistics) has been accepted by the Board of Examiners.

BOARD OF EXAMINERS


(Chairman)

Professor Dr. Tun Aung
Pro- Rector
Yangon University of Economics



(Chief Examiner)
Prof. Dr. Maw Maw Khin
Professor and Head
Department of Statistics
Yangon University of Economics



(External Examiner)
Dr. Than Pe
Lecturer (Retd.)
School of Computing and Mathematics Science
University of Western Sydney, Australia



(Examiner)
Prof. Dr. Mya Thandar
Professor
Department of Statistics
Yangon University of Economics



(External Examiner)
Daw Tin Tin Yu
Associate Professor (Retd.)
Department of Statistics
Yangon University of Economics



(External Examiner)
Daw Aye Aye Than
Associate Professor (Retd.)
Department of Statistics
Yangon University of Economics



MAY, 2018

Than Soe Oo
Head of Department
Academic Affairs
Yangon University of Economics

ABSTRACT

Infant mortality rate is considered as sensitive indicators of living standard and socio-economic conditions of a country. This study discusses the interdependency and relationship between infant mortality rate and some variables of demographic, socio-economic and housing characteristics for 54 townships in Myanmar by applying principal component analysis to determine the interdependency of these variables. The analysis produces three main components charactering each by household's education status, urban society and economic status. In order to find out the effect on infant mortality rate, multiple regression analysis technique is applied by using dominated principal components as independent variables, and infant mortality rate as dependent variable. Regression analysis explained 72.4% of total variation of infant mortality rate in Myanmar, so it is concluded that variables on household's education status, living in urban areas, and economic status in household could be main factors for determining the infant mortality rate in Myanmar.

ACKNOWLEDGEMENTS

First of all, I would like to convey my deepest gratitude to Professor Dr. Tun Aung, Pro-Rector of Yangon University of Economics, for his kind permissions for preparation of this thesis.

I would like to express my sincere appreciation to Professor Dr. Maw Maw Khin, Head of the Department of Statistics, Yangon University of Economics, for her permissions to write the thesis in this field of the study, and for providing valuable advices and suggestions to improve this thesis.

I would like to mention my gratitude to Professor Dr. Mya Thandar, Department of Statistics, Yangon University of Economics, for her valuable advices and suggestions to improve this thesis.

I would like to acknowledge Dr. Than Pe, Lecturer (Rtd.), School of Computing and Mathematics Science, University of Western Sydney, Australia, Daw Tin Tin Yu, Associate Professor (Rtd.), Department of Statistics, Yangon University of Economics and Daw Aye Aye Than, Associate Professor (Rtd.), Department of Statistics, Yangon University of Economics for their valuable comments and suggestions in preparing this thesis.

In addition, I am profoundly gratefully to all teachers from the Department of Statistics, especially to my supervisor, Daw Aye Aye Maw, Lecturer, Department of Statistics, Yangon University of Economics, for her constant guidance, in filling encouragement and valuable advices throughout the preparation of this thesis.

Finally, I would like to convey my deepest gratitude to my parents for allowing and supporting me to study in the course of Master of Economics (Statistics) in Yangon University of Economics.

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF ABBREVIATIONS	viii

CHAPTER I INTRODUCTION

1.1 Rationale of the Study	1
1.2 Objective of the Study	2
1.3 Scope and Limitations of the Study	2
1.4 Method of Study	2
1.5 Organization of the Study	3

CHAPTER II BASIC MEASURES OF DEMOGRAPHIC, SOCIO-ECONOMIC AND HOUSING INDICATORS

2.1 Demographic Indicators	4
2.1.1 Infant Mortality Rate	4
2.1.2 Total Fertility Rate	4
2.1.3 Median age	5
2.1.4 Percentage of Population Living in Urban Areas	5
2.2 Housing Indicators	5
2.2.1 Percentage of Households Using an Improved Drinking Water Source	5
2.2.2 Percentage of Households Using Improved Sanitation Facilities	6
2.2.3 Percentage of Households Using Media Facilities	6
2.2.4 Percentage of Households with Communication Devices	6
2.2.5 Percentage of Households with Motor Vehicles	7

2.3	Socio- Economic Indicators	7
2.3.1	Female Labour Force Participation Rate	7
2.3.2	Male Labour Force Participation Rate	7
2.3.3	Educational Indicators	8
(1)	Percentage of Female population that are illiterate	8
(2)	Percentage of male population that are illiterate	8

CHAPTER III THEORETICAL BACKGROUND

3.1	Principal Component Analysis	9
3.2	Population Principal Components	9
3.3	Principal Components for Covariance Matrices	10
3.4	Summarizing sample variation by Principal Components	12
3.5	The Number of Principal Components	13
3.6	Testing of Adequacy of the Approach	13
3.7	Multiple Linear Regression Model	14
3.7.1	Testing for Significance	14
3.7.2	The Coefficient of Multiple Determination	15
3.7.3	The Adjusted Coefficient of Multiple Determination	16
3.7.4	Residual Analysis	16
3.8	Multicollinearity	17

CHAPTER IV PRINCIPAL COMPONENTS ANALYSIS OF INFANT MORTALITY

4.1	Measures of Variables	19
4.2	Sampling Design	19
4.3	Mean, Standard Deviation and Correlation Coefficient Among Analyzed Variables	20
4.4	The Correlation Matrix among Independent Variables	21
4.5	Principal Component Analysis (PCA) to Determine The Infant Mortality Rate	22
4.5.1	Kaiser-Meyer-Olkin (KMO) Measure	25

4.6	Application of Principal Component Analysis in Multiple Linear Regression Model to Determine The Infant Mortality Rate	26
4.6.1	The Estimated Values of Parameter of Regression Model with Infant Mortality Rate	26
4.6.2	Testing for the Overall Multiple Regression Model	28
4.6.3	Testing for Autocorrelation of Disturbances	29

CHAPTER V	CONCLUSION	30
------------------	-------------------	-----------

REFERENCES

APPENDICES

LIST OF TABLES

Table No.	Title	Page
3.1	Durbin-Watson d Test; Decision Rules	17
4.1	Mean, Standard Deviation and Correlation Coefficient	20
4.2	Correlation Matrix of Independent Variables	21
4.3	Rotated Component Matrix	23
4.4	KMO and Bartlett's Test	25
4.5	Estimated Values of Parameters in Regression Equation	27
4.6	ANOVA	28

LIST OF FIGURES

Figure No.	Title	Page
4.1	Diagram Showing Eigenvalue and Principal Component	24
4.2	Durbin-Watson Statistic	29

LIST OF ABBREVIATIONS

ANOVA	=	Analysis of Variance
ES	=	Economic status
FLFPR	=	Female labour force participation rate
HES	=	Household's education status
KMO	=	Kaiser-Meyer-Olkin
MA	=	Median age
MLFPR	=	Male labour force participation rate
PHCD	=	Percentage of households using communication device
PHID	=	Percentage of households using an improved drinking water source
PHIS	=	Percentage of households using improved sanitation facilities
PHMF	=	Percentage of households using media facilities
PHMV	=	Percentage of households with motor vehicles
POFI	=	Percentage of female population that are illiterate
POMI	=	Percentage of male population that are illiterate.
POUP	=	Percentage of the population living in urban areas
TFR	=	Total fertility rate
US	=	Urban society
VIF	=	Variance inflation factor

CHAPTER-I

INTRODUCTION

1.1 Rationale of the Study

Health is important not only for an indicator but also for an instrument of human development. Good health can alter the population growth rate that promotes development. During the past few years, the average rate of population growth in developed countries was 0.4 and population growth in developing countries was 1.3. As a result, the average rate of population growth in developing countries is greater than in developed countries. This population growth has been of major concern not only with the government but also with the community. High population growth may eventually result in insufficient food supply, depletion of key materials, overcrowding pollution and overall deterioration in the quality of life.

Health improvement often has the greatest effect on mortality. Advances in medicine and nutrition increase the likelihood that mortality will rapidly decline. Mortality is defined as the process whereby deaths occur on a population. Status of mortality is not only an important indicator of demographic situation but also socio-economic and political situation of a country. Many demographers pay attention in studying mortality particularly for its effect upon the population size, composition, and growth. Changing in the size and structure of a population is closely related to change in variables such as economic, social, political, environmental, sanitation, etc.

Mortality rates have fallen in almost all parts of the world. Globally, ten million infants and children die each year before their fifth birthday; 99 percent of these deaths occur in developing countries and malnutrition is an underlying contributor in over half of deaths.

In Myanmar, around 56,000 children under five die in each year and 43,000 of them are younger than one month. Despite improvements, the country's infant mortality rates are the highest among ASEAN member countries. Early-age mortality rates are higher in the central plains, in rural areas, among families without formal education and among children from the poorest families. The leading causes of early-age mortality including infant mortality are birth asphyxia, pneumonia, diarrhea, malaria, measles and malnutrition¹.

¹ United Nations (2014), The Millennium Development Goals Report

Many factors contributed to infant mortality are mother's level of education, availability of water supply, sanitation improvement, access to clean drinking water and other public health infrastructure and medical measures that could help reduce high rates of infant mortality. According to the findings of 2014 Myanmar Population and Housing Census (MPHC), infant mortality rates would be reduced by improving the living standard of people. Specifically, investigation of factors on the infant mortality is important because it is essential for formulating appropriate programmes and policies in order to reduce infant mortality contributing improvement of the living standard of people. Therefore, this study explores the relationship between infant mortality and demographic, socio-economic and housing characteristics in Myanmar.

1.2 Objective of the Study

The objective of the study is to determine the effect of demographic, socio-economic and housing characteristics on infant mortality in Myanmar.

1.3 Scope and Limitations of the Study

The secondary data are obtained from the township level reports of 2014 Population and Housing Census. The study period is emphasized on the year 2014. The principal component analysis has been applied on selected demographic, socio-economic and housing variables which are mostly related to the infant mortality indicating in the past studies. The analysis is done on the administratively defined geographical areas so that findings of the study will be reflected only on those geographical areas.

1.4 Method of Study

In Myanmar, there are 412 townships under 15 states and regions including national territories. Principal component analysis has been used in order to measure the interdependency for infant mortality with demographic, socio-economic and housing characteristics in Myanmar. Principal factors accounted for high scores of variation are selected in order to compute the multiple linear regression of infant mortality on selected principal factors and to explain the most influential variables for determining infant mortality in Myanmar. Due to the unavailability of data at some

1.4 Method of Study

In Myanmar, there are 412 townships under 15 states and regions including national territories. Principal component analysis has been used in order to measure the interdependency for infant mortality with demographic, socio-economic and housing characteristics in Myanmar. Principal factors accounted for high scores of variation are selected in order to compute the multiple linear regression of infant mortality on selected principal factors and to explain the most influential variables for determining infant mortality in Myanmar. Due to the unavailability of data at some township level, the study is done on the selected townships, chosen by two- stage sample random sampling method.

1.5 Organization of the Study

This thesis is composed of five chapters. Chapter I is the introduction which includes rationale of the study, objective of the study, scope and limitations of the study, method of study and organization of the study. Chapter II presents on basic measures of demographic, socio-economic and housing indicators. Chapter III describes the theoretical background of principal component analysis and multiple linear regression analysis. Chapter IV is on the application of principal component analysis for the data computing the multiple linear regression of infant mortality on selected principal components to explain the most influential variables on infant mortality in Myanmar. The conclusion of the study with major findings and suggestions are mentioned in the Chapter V.

CHAPTER II

BASIC MEASURES OF DEMOGRAPHIC, SOCIO-ECONOMIC AND HOUSING INDICATORS

In this chapter, early-age mortality indicator such as infant mortality rate has been described, and 12 variables are chosen as the demographic, socio-economic and housing characteristics variables. These 12 variables are grouped into three categories such as demographic indicators, housing indicators, socio-economic indicators, and each variable is explained in below:

2.1 Demographic Indicators

2.1.1 Infant Mortality Rate (IMR)

The IMR is defined as the number of deaths which occurred to infants under one year of age during a given year per 1000 live births in that year.

In symbols,

$$IMR = \frac{D_0}{B} \times k$$

Where

D_0 = the number of deaths to infants under one year of age.

k = a constant

Infant mortality rate is a good indicator of reproductive and general health status of a given area or population since it tends to reflect the population's socio-economic conditions and the status of the health services.

2.1.2 Total Fertility Rate (TFR)

The total fertility rate is a summary index for age-specific fertility rate. It is calculated as the sum of a given set of age specific rates, usually those of a particular time period. It can be interpreted as measuring the number of live births a hypothetical female would have if women survived to the end of reproductive period and experienced the given set of age-specific fertility rates.

$$TFR = n \sum {}_n f_x$$

where

${}_n f_x$ = age-specific fertility rate

2.1.3 Median Age (MA)

The median age is commonly used as a basis of describing a population as young or old as ageing or growing younger. The median, a measure of central tendency is considered as the most appropriate measure of determining the average age of population. The risk of reducing infant mortality will depend on the age of mother or the average age of population. The risk of infant deaths is expected to be higher for very young mothers. The causes for high risk of infant mortality at the younger age of mothers are associated with social and physiological and reproduction immaturity. Girls who give their first birth in adolescents have poor chance of baby survival in first five years of age than those who give birth in between 20-39 years of age.

2.1.4 Percentage of the Population Living in Urban Areas (POUP)

Percentage of the population living in urban areas is also important determinants of infant mortality. Infant mortality rates in rural areas are greater than those in urban areas. This variation occurred due to unequal distribution of socio-economic factors and health facilities, for examples, the distribution of health resources in rural areas are more likely limited than in urban areas. Because of the lack of modernization and limited health facilities, rural areas are expected to have higher risk of infant mortality.

Percentage of population living in urban areas is expressed as the ratio of population living in urban areas and total population. It can be expressed as:

$$\text{POUP} = \frac{\text{Population living in urban areas}}{\text{Total population}} \times 100$$

2.2 Housing Indicators

2.2.1 Percentage of Households Using an Improved Drinking Water Source (PHID)

The Source of drinking water can cause infant mortality. Better water supply is important for child survival. Safe drinking water reduces the incidence of disease and infection which is highly related to infant mortality. Moreover, disparities in the use of improved drinking water source are also found with respect to a household's wealth level.

Percentage of households using an improved drinking water source can be calculated as households using improved drinking water source by total household.

$$\text{PHID} = \frac{\text{Households using an improved drinking water source}}{\text{Total household}} \times 100$$

2.2.2 Percentage of Households Using Improved Sanitation Facilities (PHIS)

Sanitation facilities can affect on infant mortality. It is found that the children who live in the household with improved toilets have less early-age mortality risk compared with the children who live in the house with no improved toilets. The use of improved toilets is much higher in urban and richer households compared with rural and poorer households.

Percentage of households using improved sanitation facilities is the ratio of households using improved sanitation facilities to total household.

$$\text{PHIS} = \frac{\text{Households using improved sanitation facilities}}{\text{Total household}} \times 100$$

2.2.3 Percentage of Households Using Media Facilities (PHMF)

Mother who usually uses media facilities is less likely to experience early-age death relative to the mother who does not usually use media facilities. It is assumed that high percent of possessing media facilities had low infant mortality rate.

Percentage of households using media facilities is expressed as the ratio of households using media facilities to total household.

$$\text{PHMF} = \frac{\text{Households using media facilities}}{\text{Total household}} \times 100$$

2.2.4 Percentage of Households with Communication Devices (PHCD)

Percentage of households with communication devices can be considered as an indicator of the degree of development of townships. It may also be an indicator of township's under-development and isolation. Access to communication devices indicates the socio-economic position of the household. Households in isolated and

remote areas probably have limited access to communication devices because the households may be either poor or located in under-developed areas.

Percentage of households with communication devices can be expressed as the ratio of households with communication devices to total household.

$$\text{PHCD} = \frac{\text{Households with communication devices}}{\text{Total household}} \times 100$$

2.2.5 Percentage of Households with Motor Vehicles (PHMV)

The effect of households with motor vehicles on early-age mortality appears to be similar to the availability of communication devices. It indicates the level of socio-economic development of the township. The children who live in the households with motor vehicles have less early-age mortality risk compared with the children who live in the households without motor vehicles.

Percentage of households with motor vehicles can be computed as households with motor vehicles by total household.

$$\text{PHMV} = \frac{\text{Households with motor vehicles}}{\text{Total household}} \times 100$$

2.3 Socio-Economic Indicators

2.3.1 Female Labour Force Participation Rate (FLFPR)

Female labour force participation rate is the determinant of infant mortality. Female labour force participation rate influences the income accruing to the household and the standard of living of the household. Female labour force participation rate impacts on the health and growth of the infant children. Female labour force participation rate represents the number of female in the labour force aged 15 and over per 100 population of the same aged group. It can be expressed as follows:

$$\text{FLFPR} = \frac{\text{Number of females age 15 and over in the labour force}}{\text{Total population aged 15 and over}} \times 100$$

2.3.2 Male Labour Force Participation Rate (MLFPR)

Male labour force participation rate determines the amount of income that accrues to the household which in turn affects the well-being of the infant children.

MLFPR and early-age mortality were directly related, since families with higher income had a lower infant mortality compared to the families with lower income. Male labour force participation rate represents the number of male in the labour force aged 15 and over per 100 populations of the same aged group. It can be expressed as follows:

$$\text{MLFPR} = \frac{\text{Number of males age 15 and over in the labour force}}{\text{Total population aged 15 and over}} \times 100$$

2.3.3 Educational Indicators

Educational indicators are

- (1) Percentage of female population that are illiterate (POFI) and
- (2) Percentage of male population that are illiterate (POMI).

All early-age mortality rates decrease uniformly as parental education increases. Illiteracy and low education levels of parents are associated with high mortality of infants children. Father and mother's education was significantly related to infant mortality. Parental education had an impact on child health through higher family earnings, frequent media exposure, reading and writing capability, superior health information, mother's employment status and mother's empowerment in her husband's house. Mother's education is more important than father's education. Mother's education exposes her to information about better nutrition, use of contraceptives to space births and knowledge about childhood illness and treatment.

Percentage of female population that are illiterate can be calculated as female population that are illiterate by total population.

$$\text{POFI} = \frac{\text{Female population that are illiterate}}{\text{Total population}} \times 100$$

Percentage of male population that are illiterate can be calculated as male population that are illiterate by total population.

$$\text{POMI} = \frac{\text{male population that are illiterate}}{\text{Total population}} \times 10$$

CHAPTER- III

THEORETICAL BACKGROUND

3.1 Principal Component Analysis

Principal Component Analysis (PCA) is probably the most popular multivariate statistical technique and it is used by almost all scientific disciplines. A principal component analysis is concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables. The goals of PCA are:

- (1) To extract the most important information from the data table.
- (2) To compress the size of the data set by keeping only important information.
- (3) To simplify the description of the data set; and
- (4) To analyze the structure of the observations and the variables

In order to achieve these goals, PCA computes new variables called principal components which are obtained as linear combinations of the original variables. The values of new variables for the observations are called factor scores.

Although p components are required to reproduce the total system variability often much of this variability can be accounted for by a small number of k of the principal components. There is (almost) as much information in the k components as there is in the original p variables. The k principal components can replace the initial p variables, and the original data set, consisting of n measurements on p variables, is reduced to a data set consisting of n measurements on k principal components.

3.2 Population Principal Components

Principal components are particular linear combinations of p random variables $X_1, X_2, \dots, X_p$. These linear combinations represent the selection of a new coordinate system obtained by rotating the original system with $X_1, X_2, \dots, X_p$ as the coordinate axes. The new axes represent the directions with maximum variability and provide a simpler and more parsimonious description of the covariance structure.

Principal components depend solely on the covariance matrix Σ (or the correlation matrix ρ) of $X_1, X_2, \dots, X_p$. Their development does not require a multivariate normal assumption. On the other hand, principal component derived for

multivariate normal population have useful interpretations in terms of the constant density ellipsoids.

Let the random vector $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ have the covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

The linear combinations are

$$Y_1 = \mathbf{a}_1' \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p$$

$$Y_2 = \mathbf{a}_2' \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

.

.

.

$$Y_p = \mathbf{a}_p' \mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p$$

Then,

$$\text{Var}(Y_i) = \mathbf{a}_i' \Sigma \mathbf{a}_i, \quad i = 1, 2, \dots, p$$

$$\text{Cov}(Y_i, Y_k) = \mathbf{a}_i' \Sigma \mathbf{a}_k, \quad i, k = 1, 2, \dots, p$$

The principal components are those uncorrelated linear combination $Y_1, Y_2, \dots, Y_p$ whose variances are as large as possible.

The first principal component is the linear combination with maximum variance. It maximizes $\text{Var}(Y_1) = \mathbf{a}_1' \Sigma \mathbf{a}_1$. It is clear that $\text{Var}(Y_1) = \mathbf{a}_1' \Sigma \mathbf{a}_1$ can be increased by multiplying any $\mathbf{a}_1$ by some constants. To eliminate this indeterminacy, it is convenient to restrict attention to coefficient vectors of unit length. Therefore,

First principal component = linear combination $\mathbf{a}_1' \mathbf{X}$ that maximizes $\text{Var}(\mathbf{a}_1' \mathbf{X})$
subject to $\mathbf{a}_1' \mathbf{a}_1 = 1$

Second principal component = linear combination $\mathbf{a}_2' \mathbf{X}$ that maximizes $\text{Var}(\mathbf{a}_2' \mathbf{X})$
subject to $\mathbf{a}_2' \mathbf{a}_2 = 1$ and $\text{Cov}(\mathbf{a}_1' \mathbf{X}, \mathbf{a}_2' \mathbf{X}) = 0$

i^{th} principal component = linear combination $\mathbf{a}_i' \mathbf{X}$ that maximize $\text{Var}(\mathbf{a}_i' \mathbf{X})$
subject to $\mathbf{a}_i' \mathbf{a}_i = 1$ and $\text{Cov}(\mathbf{a}_i' \mathbf{X}, \mathbf{a}_k' \mathbf{X}) = 0$ for $k < i$

3.3 Principal Components for Covariance Matrices

There are certain patterned covariance and correlation matrices whose principal components can be expressed in simple form. Suppose Σ is the diagonal matrix

$$\Sigma = \begin{bmatrix} \sigma_{11} & 0 \dots & 0 \\ 0 & \sigma_{22} \dots & 0 \\ 0 & 0 \dots & \sigma_{pp} \end{bmatrix}$$

Setting $e_i = [0, \dots, 0, 1, 0, \dots, 0]$, with 1 in the i^{th} position ,

$$\begin{bmatrix} \sigma_{11} & 0 \dots & 0 \\ 0 & \sigma_{22} \dots & 0 \\ 0 & 0 \dots & \sigma_{pp} \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ \sigma_{ii} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$(\text{or}) \Sigma e_i = \sigma_{ii} e_i$$

and its conclude that (σ_{ii}, e_i) is the i^{th} eigenvalue-eigenvector pair. Since the linear combination $e_i' X = X_i$, the set of principal components is just the original set uncorrelated random variables.

Another patterned covariance matrix, which often describes the correspondence among certain biological variables.

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \dots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \dots & \rho\sigma^2 \\ \vdots & \vdots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

The resulting correlation matrix

$$\rho = \begin{bmatrix} 1 & \rho \dots & \rho \\ \rho & 1 \dots & \rho \\ \vdots & \vdots & \vdots \\ \rho & \rho & 1 \end{bmatrix}$$

is also the covariance matrix of the standardized variables. The matrix implies that the variables $X_1, X_2, \dots, X_p$ are equally correlated.

The p eigenvalues of the correlation matrix can be divided into two groups. When ρ is positive, the largest is

$$\lambda_1 = 1 + (p-1) \rho$$

with associated eigenvector

$$\mathbf{e}_1' = \left[\frac{1}{\sqrt{p}}, \frac{1}{\sqrt{p}}, \dots, \frac{1}{\sqrt{p}} \right]$$

The remaining $p-1$ eigenvalues are

$$\lambda_2 = \lambda_3 = \dots = \lambda_p = 1 - \rho$$

and one choice for their eigenvectors is

$$\mathbf{e}_2' = \left[\frac{1}{\sqrt{1 \times 2}}, \frac{-1}{\sqrt{1 \times 2}}, 0, \dots, 0 \right]$$

$$\mathbf{e}_3' = \left[\frac{1}{\sqrt{2 \times 3}}, \frac{1}{\sqrt{2 \times 3}}, \frac{-2}{\sqrt{2 \times 3}}, 0, \dots, 0 \right]$$

$\vdots$

$$\mathbf{e}_i' = \left[\frac{1}{\sqrt{(i-1)i}}, \dots, \frac{1}{\sqrt{(i-1)i}}, \frac{-(i-1)}{\sqrt{(i-1)i}}, 0, \dots, 0 \right]$$

$\vdots$

$$\mathbf{e}_p' = \left[\frac{1}{\sqrt{(p-1)p}}, \dots, \frac{1}{\sqrt{(p-1)p}}, \frac{-(p-1)}{\sqrt{(p-1)p}} \right]$$

The first principal component

$$Y_1 = \mathbf{e}_1' \mathbf{Z} = \frac{1}{\sqrt{p}} \sum_{i=1}^p Z_i$$

is proportional to the sum of the p standardized variables. It might be regarded as an “index” with equal weights. This principal component explains a proportion

$$\frac{\lambda_1}{p} = \frac{1 + (p-1)\rho}{p} = \rho + \frac{1-\rho}{p}$$

of the total population variation. It can be seen that $\frac{\lambda_1}{p} = \rho$ for ρ close to 1 or p large.

3.4 Summarizing Sample Variation by Principal Components

The sample principal components are defined as those linear combinations which have maximum sample variance. As with the population quantities, the coefficient vectors $\mathbf{a}_i$ to satisfy $\mathbf{a}_i' \mathbf{a}_i = 1$

First *sample* principal component = linear combination $\mathbf{a}_1' \mathbf{X}_j$ that maximizes the sample variance of $\mathbf{a}_1' \mathbf{X}_j$ subject to $\mathbf{a}_1' \mathbf{a}_1 = 1$

Second *sample* principal component = linear combination $\mathbf{a}_2' \mathbf{X}_j$ that maximizes the sample variance of $\mathbf{a}_2' \mathbf{X}_j$ subject to $\mathbf{a}_2' \mathbf{a}_2 = 1$ and zero sample covariance for the pairs $(\mathbf{a}_1' \mathbf{X}_j, \mathbf{a}_2' \mathbf{X}_j)$

i^{th} sample principal component = linear combination $\mathbf{a}_i' \mathbf{X}_j$ that maximizes the sample variance of $\mathbf{a}_i' \mathbf{X}_j$ subject to $\mathbf{a}_i' \mathbf{a}_i = 1$ and zero sample covariance for all pairs $(\mathbf{a}_i' \mathbf{X}_j, \mathbf{a}_k' \mathbf{X}_j)$, $k < i$

3.5 The Number of Principal Components

A useful visual aid to determining an appropriate number of principal components is a *scree plot*. With the eigenvalues ordered from largest to smallest, a scree plot is a plot of $\hat{\lambda}_i$ versus i - the magnitudes of an eigenvalue versus its number. To determine the appropriate number of components, an elbow (bend) in the scree plot is important. The number of components is taken to be the point at which the remaining eigenvalues are relatively small and all about the same size.

3.6 Testing of Adequacy of the Approach

Kaiser-Meyer-Olkin (KMO) and Bartlett's Test of sphericity is a measure of sampling adequacy that is recommended to check the case to variable ratio for the analysis being conducted. KMO and Bartlett's test play an important role for accepting the sample adequacy.

For a large sample, Bartlett's test approximates a chi-square distribution. Bartlett's test of sphericity tests the hypothesis that correlation matrix is an identity matrix, which would indicate that variables are unrelated and therefore unsuitable for structure detection. For principal component analysis to be recommended suitable, the Bartlett's test of sphericity must be less than 0.05.

Kaiser-Meyer-Olkin measure of sample adequacy compares magnitudes of observed correlation coefficients to partial correlation coefficients. The KMO ranges from 0 to 1. KMO values between 0.8 and 1 indicate the sampling is adequate. KMO values less than 0.6 indicate the sampling is not adequate and that remedial action should be taken. KMO values close to zero means that there are large partial correlations compared to the sum of correlation. Kaiser gives values of

- KMO > 0.9 marvelous
- 0.8 < KMO < 0.9 meritorious
- 0.7 < KMO < 0.8 middling
- 0.6 < KMO < 0.7 mediocre
- 0.5 < KMO < 0.6 miserable
- KMO < 0.5 unacceptable

3.7 Multiple Linear Regression Model

Multiple regression analysis is one of the most widely used of all statistical tools to describe simultaneously a statistical relation among variables known as independent and dependent variables. A multiple regression is used to examine the relationship between one dependent variable and a set of independent variables when two or more independent variables are used in making estimates of the dependent variable. The equation which describes how the dependent variable Y is related to the independent variables $X_1, X_2, \dots, X_k$ and an error term e is called the multiple regression model.

Multiple regression models take the following form;

$$Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ik} + e_i; i=1, 2, \dots, n$$

Where;

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

β_0 = constant in the regression equation

$\beta_1, \dots, \beta_k$ = regression coefficients associated with each of the independent variable.

X_{ij} = value of the j^{th} independent variable in the i^{th} trial

e_i = the random error in the i^{th} trial or observation

The value of $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ are unknown and will have to estimate them from sample data. Using this sample, an estimated multiple regression can develop which takes the following form;

$$\hat{Y}_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \dots + b_k X_{ik}$$

Where $b_0, b_1, b_2, \dots, b_k$ are the estimated values for the parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ and $\hat{Y}_i$ is the estimated value of the dependent variable.

3.7.1 Testing For Significance

The significance tests for the simple regression model were the t test and F test. In multiple regression analysis, the t test and F test have different purposes.

1. 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.

2. 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.

Test for the Overall Multiple Regression Model

To test whether there is a regression relation between the response variable Y and the set of X variables $X_1, \dots, X_k$, i.e., to choose between the alternatives;

Null Hypothesis $H_0; \beta_1 = \dots = \beta_k = 0$

Alternative Hypothesis H_a ; not all $\beta_i (i=1, \dots, k)$ equal zero

The ratio of test statistics;

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

The decision rule for this test,

If $F^* > F(1-\alpha, k, n-k-1)$, Reject H_0

If $F^* \leq F(1-\alpha, k, n-k-1)$, Do not reject H_0

The existence of a regression relation by itself does not ensure that useful predictions can be made by using it.

Test for the Individual Regression Coefficient

An individual partial regression coefficient, β_j in the multiple regression model is tested to determine the significance of the relationship between X_j 's and Y .

Null Hypothesis $H_0; \beta_j = 0$

Alternative Hypothesis $H_a; \beta_j \neq 0$

The test statistic is

$$t^* = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

The decision rule for this test,

If $t^* > t(1-\alpha/2, n-k-1)$, Reject H_0

If $t^* \leq t(1-\alpha/2, n-k-1)$, Do not reject H_0

3.7.2 The Coefficient of Multiple Determination R^2

The coefficient of multiple determination, denoted by R^2 , is defined as follows:

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}$$

The coefficient of multiple determination, what proportion of the total variability in Y, the dependent variables, is explained by the independent variables. That is the percentages of the total variation of the dependent variables that can be explained by the explanatory variable. The value of R^2 will be between zero and one, where $R^2 = 0$, the regression model cannot explain anything about the variation in the dependent variable or the estimated model does not fit the data. The case of $R^2 = 1$ represents a perfect fit of the estimated model of the data. A high value of R^2 shows good fit and a low value of R^2 shows a poor fit.

3.7.3 The Adjusted Coefficient of Multiple Determination $\bar{R}^2$

A measure that recognized the number of independent variables in the regression model is called the adjusted coefficient of multiple determination and is denoted by $\bar{R}^2$.

$$\bar{R}^2 = \frac{\frac{\sum(Y_i - \bar{Y})^2}{(n-k-1)}}{\frac{\sum(Y_i - \bar{Y})^2}{(n-1)}}$$

Reporting the adjusted R^2 is extremely important in comparing two or more regression models that predict the same dependent variable but have a different number of independent variables.

3.7.4 Residual Analysis

The Durbin-Watson Statistic is used to detect autocorrelation. The following equation defines the Durbin-Watson statistic.

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

e_i = residual at time period

As a general rule, if d is close to 2, assume that autocorrelation is not a problem.

Table (3.1)**Durbin-Watson d test: Decision Rules**

Null hypothesis	Decision	If
No positive autocorrelation	Reject	$0 < d < d_L$
No positive autocorrelation	No decision	$d_L \leq d \leq d_U$
No negative correlation	Reject	$4 - d_L < d < 4$
No negative correlation	No decision	$4 - d_U \leq d \leq 4 - d_L$
No autocorrelation, positive or negative	Do not reject	$d_U < d < 4 - d_U$

3.8 Multicollinearity

Multicollinearity problem 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 conditions for multiple regressions. Specifically, multicollinearity occurs if there is a high correlation between two independent variables, X_i and X_j if the correlation coefficient r_{ij} between X_i and X_j in the multiple linear regression model is high, multicollinearity exist. Multicollinearity is a problem of degree. Any time two or more independent variables are linearly related, some degree of multicollinearity exists. If its presence becomes too pronounced, the model is adversely affected. The presence of multicollinearity creates many problems in use of multiple linear regression models.

The most direct way of testing for multicollinearity is to produce a correlation matrix for all variables in the model. Another way to detect multicollinearity is to use the value of Tolerance. If the value of Tolerance is not less than 0.1, it can be said that there is no multicollinearity problem in this study. The another way to detect multicollinearity is to use the variance inflation factor (VIF). The VIF associated with any x variable is found by regression it on all the other X variables. The resulting R^2 is then used to calculate that variable's VIF. The VIF for any independent variable is a measure of the degree of the multicollinearity contributed by that variable.

The VIF for any independent variable X_i is

$$VIF(X_i) = \frac{1}{1 - R^2}$$

Where, R^2 is the coefficient of determination obtained by regression X_i on all other independent variables. Multicollinearity produces an increase in the variation or standard error of the regression coefficient. VIF measures the increase in the variance of the regression coefficient over that which would occur if multicollinearity were not present. In general, multicollinearity is not considered a significant problem unless the VIF of a single X_i measure at least 10 or the sum of the VIF's for all X_i is at least 10.

CHAPTER- IV

PRINCIPAL COMPONENT ANALYSIS OF INFANT MORTALITY RATE

The information for analysis is taken from the township level reports of 2014 Population and Housing Census in Myanmar to calculate a principal component analysis, and multiple regression analysis. Detailed Data relating to the variables analyzed in the township level is mentioned in the **Appendix –A**.

4.1 Measures of Variables

The analysis is done with a group of variables consisting of 13 variables which are classified into one dependent variable and 12 independent variables. Dependent variable is infant mortality rate (IMR) and the independent variables are percentages of households using an improved drinking water source (PHID), percentages of households using improved sanitation facilities (PHIS), percentages of households using media facilities (PHMF), percentages of households using communication devices (PHCD), percentages of households with motor vehicles (PHMV), female labour force participation rate (FLFPR), male labour force participation rate (MLFPR), percentage of female population that are illiterate (POFI), percentage of male population that are illiterate (POMI), total fertility rate (TFR), median age (MA) and percentage of the population living in urban areas (POUP).

4.2 Sampling Design

In Myanmar, there are 412 townships in 15 States and Regions including national territories. Since there are not available the township level reports at time of preparation of this thesis, the sampling design of the selection of Townships is chosen of two- stage random sampling method were chosen in order to be more convenience of selecting sample units . The first stage of sampling unit is choosing the state and regions by applying simple random sampling method. Out of 15 states and regions, 7 states and regions were fallen in first stage sample. Those states and regions are Yangon Region, Sagaing Region, Mandalay Region, Ayeyarwaddy Region and Kachin, Rakhine, Shan States.

The second stage sampled unit is townships from the first stage sample. Applying simple random sampling again, a total of 54 townships were selected from the first stage sample of state and regions. Those sampled townships were mentioned in **Appendix- A**.

4.3 Mean, Standard Deviation and Correlation Coefficient among Analyzed Variables

Table (4.1) states the descriptive measures and correlation coefficients of the demographic, socio- economic and housing characteristics variables.

Table (4.1)
Mean, Standard Deviation and Correlation Coefficient

Independent Variables	Mean	Standard Deviation	Pearson Correlation Coefficient
PHID	65.2407	29.07394	-0.667**
PHIS	72.2981	20.91143	-0.544**
PHMF	53.4556	23.60305	-0.815**
PHCD	38.0593	30.09669	-0.849**
PHMV	43.9515	22.94383	-0.125
FLFPR	55.8241	13.42307	0.061
MLFPR	85.3574	5.19730	-0.569**
POFI	18.2685	19.93259	0.136
POMI	10.3037	14.13620	0.077
TFR	2.4772	0.88515	0.623**
MA	27.1685	3.10032	-0.436**
POUP	35.4796	36.09254	-0.740**

Source: SPSS Output

Note: ** = Correlation is significant at the 0.01 level.

* = Correlation is significant at the 0.05 level.

According to Table (4.1), strong significant negative correlation is found between infant mortality rate and percentage of households using media facilities

(PHMF), percentage of households using communication devices (PHCD) and percentage of the population living in urban areas (POUP). Fair significant negative correlation is found between infant mortality rate and percentage of households using an improved drinking water source (PHID), percentage of households using improved sanitation facilities (PHIS), male labour force participation rate (MLFPR) and median age (MA) while percentage of households with motor vehicles (PHMV) have weak negative relationship with infant mortality rate. However, female labour force participation rate (FLFPR), percentage of female population that are illiterate (POFI) and percentage of male population that are illiterate (POMI) has weak positive relationship with infant mortality rate while total fertility rate (TFR) have fair significant positive relationship with infant mortality rate.

4.4 The Correlation Matrix among Independent Variables

The correlation matrix among the independent variables is shown in Table (4.2).

Table (4.2)

Correlation Matrix of Independent Variables

Variable	PHID	PHIS	PHMF	PHCD	PHMV	FLFPR	MLFPR	POFI	POMI	TFR	MA	POUP
PHID	1.000											
PHIS	.725	1.000										
PHMF	.629	.710	1.000									
PHCD	.623	.644	.950	1.000								
PHMV	.134	-.011	.025	-.127	1.000							
FLFPR	-.022	.008	-.037	-.148	.498	1.000						
MLFPR	-.375	-.418	-.603	-.712	.365	.538	1.000					
POFI	-.317	-.503	-.147	-.209	.542	.322	.378	1.000				
POMI	-.286	-.462	-.072	-.127	.515	.312	.335	.971	1.000			
TFR	-.677	-.685	-.638	-.673	.146	.063	.386	.513	.491	1.000		
MA	.629	.641	.483	.516	-.195	-.132	-.353	-.446	-.464	-.838	1.00	
POUP	.527	.588	.884	.935	-.198	-.142	-.774	-.260	-.191	-.570	.435	1.000

According to the Table (4.2), there is high correlation between percentages of households using an improved drinking water source (PHID) with percentage of households using improved sanitation facilities (PHIS). There are high correlation between percentage of households using improved sanitation facilities (PHIS) and percentage of households using media facilities (PHMF). High correlation coefficients are seen percentage of households using media facilities (PHMF) with percentage of

households using communication devices (PHCD) and percentage of the population living in urban areas (POUP). There are high correlation among percentage of households using communication devices (PHCD) with male labour force participation rate (MLFPR) and percentage of population living in urban areas (POUP). There are moderate correlation among percentage of households with motor vehicles (PHMV) with female labour force participation rate (FLFPR), male labour force participation rate (MLFPR), percentage of female population that are illiterate (POFI) and percentage of male population that are illiterate (POMI). There are high correlation between male labour force participation rate (MLFPR) and percentage of population living in urban areas (POUP). Percentage of female population that are illiterate (POFI) and percentage of male population that are illiterate (POMI) has strong relationship. There are moderate correlation among percentage of male population that are illiterate (POMI), total fertility rate (TFR) and median age (MA). There are high correlation between total fertility rate (TFR) and median age (MA). Median Age (MA) is fairly associated with percentage of population living in urban areas (POUP). There are fair correlation of female labour force participation rate (FLFPR) with male labour force participation rate (MLFPR), percentage of female population that are illiterate (POFI) and percentage of male population that are illiterate (POMI).

Table (4.1) and Table (4.2) show that independent variables are highly correlated with the dependent variables in which they themselves are highly correlated with one another concluding that there may cause multicollinearity problems among variables to be considered in multiple linear regression models..

4.5 Principal Component Analysis (PCA) to Determine the Infant Mortality Rate

The principal component analysis (PCA) is employed to find out the effects among independent variables before finding the determinants of infant mortality. The principal component using Principal Component Analysis (PCA) is conducted using Statistical Packages for Social Sciences (SPSS). The results of all principal components from the data are mentioned in **Appendix-B**. Table (4.3) below shows the results of the rotated principal component analysis.

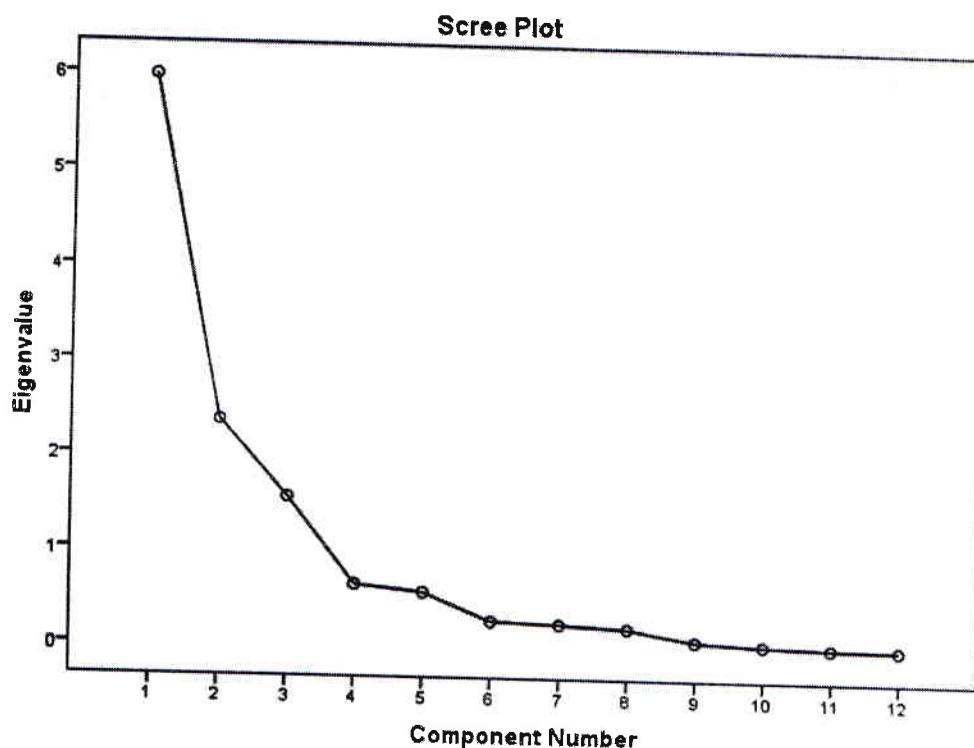
Table (4.3)
Rotated Component Matrix

Variables	Component		
	1	2	3
PHCD	.948	.234	-.075
PHMF	.929	.247	.082
POUP	.913	.184	-.180
MLFPR	-.330	-.089	.562
POMI	.104	-.784	.243
POFI	.029	-.770	.258
TFR	-.496	-.756	-.002
MA	.378	.747	-.033
PHIS	.539	.705	.085
PHID	.561	.609	.207
PHMV	.020	-.161	.837
FLFPR	-.155	.060	.799
Eigenvalues	5.982	2.366	1.558
% of variance	49.849	19.719	12.985
Cumulative %	49.849	69.568	82.554

Source: SPSS Output

Table (4.3) displays eigenvalues, percent and cumulative percent of explained variance which helps us to decide how many components are being retained. By considering scree plot in Figure (4.1), it is reasonable to retain first three components as these eigenvalue is more than 1.

Figure (4.1) Diagram Showing Eigenvalue and Principal Component



Source: SPSS Output

It can be seen in Table (4.3) that the principal component analysis produced three components whose eigenvalues are greater than 1 and these components explain more than 82.6% of total variation of all variables.

The Table shows that 49.8% of total variation of sample is explained by the first principal component, 19.7% of variation by second component and 12.9% of variation by third component. So, the first three principal components explain 82.6% of the total sample variance. The cumulative proportion of variance criteria meet three principal components to satisfy the criterion of explaining 82.6% of the total variance, leaving 17.4 % of variation to other unexplained factors.

Component 1 having an eigenvalue of 5.982, explains 49.849% of total variance of the PCA. Component 1 can be said influencing by a set of variables of PHID (percentage of households using an improved drinking water source, PHIS (percentage of households using improved sanitation facilities), PHMF (percentage of households using media facilities), PHCD (percentage of households using

communication devices) and POUP (percentage of the population living in urban areas). Component 1 can be named as urban society (US).

Component 2 has an eigenvalue of 2.366, and explains 19.719% of the total variance of the variables analyzed. Component 2 is said to be influenced by a set of variables consisting of PHID (percentage of households using an improved drinking water source, PHIS (percentage of households using improved sanitation facilities), POFI (percentage of female population that are illiterate), POMI (percentage of male population that are illiterate), TFR (total fertility rate) and MA (median age). Component 2 can be called as household's education status (HES).

Component 3 with its eigenvalue of 1.558, explains 12.985% of the total variance. Component 3 is influenced by the variables of FLFPR (female labour force participation rate), MLFPR (male labour force participation rate) and PHMV (percentage of households with motor vehicles). Component 3 can be named as economic status (ES) of the household.

4.5.1 Kaiser-Meyer-Olkin (KMO) Measure

In order to verify the suitability of employing principal component analysis to the variables considered, the results of the KMO and Bartlett's Test of Sphericity are shown in the following Table (4.4).

Table (4.4)

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measures of Sampling Adequacy		0.774
Bartlett's Test of Sphericity	(Approx. Chi-Square)	744.952
Degree of freedom		66
P-value		0.000

Source: SPSS Output

Kaiser-Meyer-Olkin Measure of Sampling Adequacy compares magnitudes of observed correlation to partial correlation coefficients. The value of township's characteristics explained by the set of variables included in the analysis is shown in 0.774, which is middling sample adequacy for relationship between township's characteristics of the set of variables. On the other hand, the values (0.774) of Kaiser-Meyer-Olkin measure of sampling adequacy is higher than 0.5 indicating that solution through principal component analysis is suitable for the analysis of variables.

Likewise, Chi-Square (Bartlett's Test of Sphericity) value is shown as 744.952 that is explaining that null hypothesis of variables are uncorrelated is rejected at Bartlett's test of Sphericity in terms of Chi-Square test statistic, which is less than conventional value 0.05.

4.6 Application of Principal Component Analysis in Multiple Linear Regression Model to Determine the Infant Mortality Rate

Applying three principal components dominated in the variables analyzed, a sample multiple regression equation for IMR could be written as:

$$\text{IMR} = b_0 + b_1 \text{ US} + b_2 \text{ HES} + b_3 \text{ ES}$$

Where b_0 , b_1 , b_2 , b_3 are regression coefficients, and IMR (Infant Mortality Rate) as dependent variables and principal components are assumed as independent variables naming as US (Urban Society), HES(Household's Education Status) and ES (Economic Status).

4.6.1 The Estimated Values of Parameter of Regression Model with Infant Mortality Rate

The estimated parameter values for regression equation have been expressed in the following Table (4.5).

Table (4.5)
Estimated Values of Parameters in Regression Equation

Variables	Coefficient	Std. Error	t	Sig	Tolerance	VIF
(Constant)	54.519**	1.663	32.776	0.000		
Urban society	-18.302**	1.679	-10.900	0.000	1.000	1.000
Household's education status	-5.499**	1.679	-3.275	0.002	1.000	1.000
Economic status	-2.268*	1.679	-1.351	0.073	1.000	1.000
R	0.851					
R ²	0.724					
Adjusted R ²	0.708					
Durbin-Watson	1.669					

Source: SPSS Output

Note; ** = 1% significant, * = 10% significant.

According to Table (4.5), the estimated multiple regression model is

$$\text{IMR} = 54.519 - 18.302 \text{ US} - 5.499 \text{ HES} - 2.268 \text{ ES}$$

$$\text{P value} = (0.000) \quad (0.000) \quad (0.002) \quad (0.073)$$

From the estimated regression equation, it is found that household's education status (HES) and economic status (ES) are held constant; one percent increase in urban society (US) led on the average to about 18.302 percent will decrease in infant mortality rate (IMR). Similarly, it can be found that urban society (US) and economic status (ES) are held constant; one percent increase in the household's education status (HES) led on the average to about 5.499 percent will decrease in infant mortality rate (IMR). Then, holding urban society (US), household's education status (HES) are constant and economic status (ES) goes up by one percent, infant mortality rate (IMR) will decrease by 2.268 percent.

According to the Table (4.5), the regression coefficients of urban society (US) and household's education status (HES) have at 1% level of significance. The coefficient of the variable of economic status (ES) is significant at 10% level.

The regression equation produces the value of coefficient of determination; R^2 is 0.724 meaning that 72.4% of variation in infant mortality rate is explained by living in urban society (US), household's education status (HES) and economic status (ES) of household in the townships. The higher the value of R^2 , the more explanatory power the model has.

As shown in Table (4.5), the value of VIF for each explanatory variable is one. So, it can be concluded that there is no seriously multicollinearity problem in this regression model.

4.6.2 Testing for the Overall Multiple Regression Model

Table (4.6) shows the analysis of variance (ANOVA) using to test whether any of the explanatory variables has a relationship with the explained variable.

Table (4.6)
ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig
Regression	19627.127	3	6542.376	43.789	0.000
Residual	7470.354	50	149.407		
Total	27097.481	53			

Source: SPSS Output

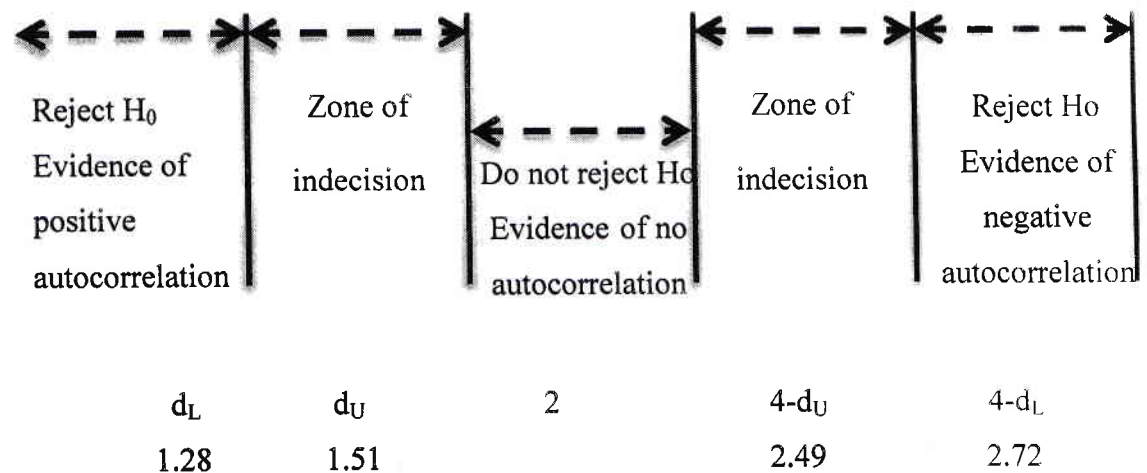
According to above Table (4.6), the value of F statistic is 43.789. It is said that to be statistically significant at 1 percent level, given the fact that the calculated probability value is 0.000. This implies that the overall model is statistically significant.

4.6.3 Testing for Autocorrelation of Disturbance

The Durbin-Watson statistics is used to test the hypothesis of no autocorrelation. As seen in Table (4.5), the Durbin-Watson test statistic is 1.669 which falls between the zone, for $\alpha = 0.01$ or 1% level of significance, critical values for the Durbin-Watson statistic (figure 4.2) are $d_U = 1.51$ and $4-d_U = 2.49$, stating that the null hypothesis is not rejected and it is concluded that there is no evidence of autocorrelation. Therefore, D.W statistic satisfied that there is no positive and negative correlation between disturbances. As a general rule, if d is close to 2, it is assumed that autocorrelation is not a problem.

Figure (4.2)

Durbin-Watson Statistic



CHAPTER- V

CONCLUSION

In this thesis, the study is done by employing principal component analysis to determine the effect of infant mortality on selected variables related to demographic, socio-economic and housing characteristics in Myanmar. According to 2014 Myanmar Population and Housing Census, there are around 51 million population in Myanmar. 30% of peoples are live in the urban areas and the remaining peoples live in the rural areas. It is observed that continued declining trend of infant mortality in Myanmar but it is varied among townships in Myanmar due to unequal distribution of socio-economic conditions, health infrastructures and other variables associated with environmental and sanitation factors, housing condition, educational status, and economic status households in townships.

This study is focused on the effect of demographic, socio-economic and housing characteristics on infant mortality in Myanmar by using Principal Component Analysis (PCA) and Multiple Regression Analysis, which was carried out using Statistical Packages for Social Sciences (SPSS).

Different levels of associations among variables can be seen. Strong significant negative correlation is found between infant mortality rates (IMR) and PHMF, PHCD, POUP, PHID, PHIS, MLFPR, MA. However, there are also observed that positive relationship with infant mortality rate (IMR) with FLFPR, POFI, POMI, and TFR.

There are common variables that are positively correlated with the principal components. It includes the variables of PHMF, PHCD, POUP, PHID, PHIS, FLFPR, MLFPR, MA, PHMV, and negatively related are included as POFI, POMI and TFR.

Calculated principal components results showed that three influenced components which account for 82.6 % total variation of the system could be considered as the determinants of infant mortality. Component 1 is named as urban society (US) is contributed 49.8%, component 2 which is called as household's education status (HES) contributing to 19.7% and component 3 is named as economic status (ES) is contributed 12.9% of total variation of the system. These three principal components could be considered as the main influenced factors of the system representing 12 variables for finding out the determinants of the variation of the infant mortality in Myanmar.

Estimation of the multiple regression model with infant mortality on 3 main principal factors have resulted as significant model construction explaining 72.4% of sample variation of infant mortality rate in Myanmar. Each component considered in the regression model is statistically significant at 1% level and 10% level. It could be interpreted that the more the status of those components, the lower the infant mortality could be with 72.4% certainty.

The suitability of using sample data in principal components analysis is measured by Kaiser-Meyer-Olkin (KMO) Measure which value is 0.774 and Bartlett's Test of Sphericity that values is 744.952, both of them were significant at 0.01 level. These test indicated that the using sample data in principal component analysis is suitable for the constructed model and it is also found that there is no relationship with independent and dependent variables considered in this study. Testing for autocorrelation of disturbances by using Durbin-Watson test statistic (the value of 1.669) is concluded that there is no evidence of autocorrelation among the disturbance used in the model.

In general, infant mortality rate is varied according to living in urban areas, household's education status and economic status in household in townships of Myanmar. The higher the values of indicators showing those variables are, the lower the infant mortality rate could be in Myanmar. It is proposed to do more investments in those three broad areas implementing programme or projects supporting to increase accessibility of media facilities, communication devices, urbanization, environmental sanitation, female and male labour force participation and having motor vehicles in households and to make effort to decrease the illiteracy of male and female and total fertility.

REFERENCES

1. Ajaero, C. K and Owch, O. L (2013), "The pattern and Determinants of Infant Mortality in Rural Nigeria, Journal of Social Science, Vol.1, Issue 1.
2. Department of Population (2014), The 2014 Myanmar Population and Housing Census (National level), Ministry of Immigration and Population, Nay Pyi Taw.
3. Department of Population (2014), Township level Population and Housing Census report, Ministry of Immigration and Population, Nay Pyi Taw.
4. Godson, C. and Nnamdi, J. (2012), " Environmental Determinants of Infant Mortality in Nigeria", Journal of Sustainable Development, Vol.5, No.1.
5. Humera, R., Mehboob, A. and Maqsood, A. (2015), "Principal Component Analysis and Socio-economic Factors and Their Association with Life Expectancy at Birth in Asia", Journal of Multidisciplinary Academic Research, Vol.3, No.1.
6. Johnson, R. A and Wichern D.W (2002), "Applied Multivariate Statistical Analysis, 6th Edition, United State of America
7. Kibet and Mases, K. (1981), "Differential Mortality in Kenya", Master Thesis, University of Nairobi.
8. Ohnmar Oo (2017), "Principal Component Analysis of Urban Population in Myanmar", Master Thesis, Yangon University of Economics.
9. Thu Thu Han Min (2014), "Solving Multicollinearity Problem by Principal Component Regression", Master Thesis, Yangon University of Economics.
10. Webster, A. L. (1995), "Applied Statistical for Business and Economics", (Second Edition), McGraw-Hill, Inc.
11. Win Myint (1988), "Multivariate Analysis of Areal Fertility in Burma", Master Thesis, Yangon Institute of Economics.

APPENDICES

APPENDIX (A)

Indicators by Selected Townships

Sr No.	Townships	IMR	PHID	PHIS	PHMF	PHWC	PHMV	FLFPR	MLFPR	POFI	POMI	TFR	MA	POUP
1	Yankin	21	99.2	99	93.1	90.7	26.87	47.5	75.6	3.3	1.3	1.26	31.9	100
2	South-Okkalapa	18	99.7	97.6	90.7	82.7	20.94	44.8	77.5	2.4	0.8	1.31	31.8	100
3	Kyauktada	19	98.4	99.8	93.7	94.9	25.52	47.1	77.6	2.9	0.9	1.27	32.6	100
4	Lanmadaw	18	99.6	99.5	94.3	95	33.6	55.8	77.4	1.8	0.7	0.94	30.5	100
5	Latha	38	99.5	100	93.1	92.6	28.88	58.6	81.1	3.2	1	0.93	31.7	100
6	Kyimyindine	46	68.3	75.1	81.1	74.8	17.19	47.7	80.6	5.2	2.2	1.76	28.9	100
7	Hline	19	99.4	99.3	92.2	89.4	20.73	50.6	79.9	2.5	0.9	1.19	30.5	100
8	Kamayut	24	99.7	99.8	91.7	95.3	27.35	54.9	76.6	1.3	0.5	1.22	29.5	100
9	Dagon	14	99.9	99.5	96.3	95.4	37.24	53.5	78.3	1.7	0.6	1.22	30.4	100
10	Bahan	18	99.5	98.4	94.2	93.7	26.74	54.2	74.3	1.9	0.6	1.22	30.9	100
11	Seikkan	26	99.8	99.8	85.4	92.7	9.76	50.4	79.9	1	0.5	2.19	28	100
12	Kangyidaunt	74	48.2	82.3	46.1	21.5	18	40.4	85.7	8.5	4.9	2.74	26.6	6.2
13	Ngaputaw	83	35.3	48.1	34.4	11.5	12	34.4	89.5	9.1	6.9	3.05	25.4	6.3
14	Thapaung	81	67.4	57.4	37.5	17.8	14.1	41	86.1	8.5	4.2	2.91	26.7	4.7
15	Phyapon	85	0.9	61.1	40	27.9	16.3	30.4	80.3	7.3	3.4	2.78	26.9	26.2
16	Wakema	83	16	79.4	39.3	16.9	10.5	49.7	87	11.5	6.6	3.11	27.5	7.7
17	Einne	80	43.3	76.1	43.6	15.1	16.9	38.1	85.8	12.1	8	2.85	27	7
18	Ingapu	78	90.7	75.3	32.8	17.8	25.4	40.3	79.7	5.7	2.1	2.41	30.6	6.2
19	Sinku	43	73.5	70.1	41.9	24.4	49.2	41.1	87	4.5	1.7	2.64	26	4.7
20	Myithar	53	87.1	74.8	51.3	28.4	53.5	51.9	90	9.9	3.1	2.26	28.5	9.9
21	Tada U	45	87.7	76.7	48.2	31.1	63.6	46.3	87.7	7.8	2	2.15	30.1	8.9
22	Kyaukpadaung	58	70	83.3	38.6	31.2	42	60.7	84.1	13.9	3.1	2.13	30.1	16.3

APPENDIX (A) (cont.)

23	Yamethin	64	86.4	79.3	33.6	22.7	44.7	53.2	90	17.1	3.8	2.21	28.4	10.7
24	Wundwin	54	91.9	80.5	39.8	24.4	56.5	64.8	89.7	13.4	4.3	1.98	30.4	14.2
25	Myingyan	47	82.3	70.4	43.9	33.1	56.8	64.6	83.8	9.7	3.1	1.92	30.2	31.8
26	Wetlet	49	66.7	60.7	34.2	16	49.9	64.2	90.2	8.2	2.3	1.90	31.3	5.5
27	Kambalu	52	76.5	65	32.6	14.3	51.5	69.7	88.4	5.8	1.5	2.45	26.1	8.5
28	Ayartaw	50	86.3	67.8	28.2	13.5	61.7	44.2	88.4	13.3	4.1	1.94	30.3	6.1
29	Kalewa	84	47.6	86	39	15.9	38.6	80.2	93.4	1.1	0.5	2.38	26.8	15.6
30	Mawlaik	85	77.6	90.3	46.2	17.8	38.9	69	92.5	2.2	0.9	2.63	25.9	16.7
31	Phaungpyin	76	84.4	76.6	30.1	1.8	42.9	74.5	92	4.4	1.2	3.55	22.6	8.9
32	Salingyi	60	87.2	72.1	44.4	21.8	61.2	58.7	85.6	8.5	2.9	1.86	30.1	4.6
33	Myitkyina	41	92.2	83.5	68.8	56.3	85	39.7	75.6	7.7	3.5	2.69	24.5	79.2
34	Waingmaw	62	67.9	84	46.8	23.6	65.5	45.7	83.2	20.6	12.6	3.39	23.1	20.7
35	Sumprabum	66	10.4	50.9	34.7	9.4	45.5	53.3	81.1	29.8	16.6	5.61	22.1	44.7
36	Myauk U	72	13.3	20.2	17.5	9.8	8.7	39.8	88.4	19.1	6.7	2.66	24.4	19.1
37	Minbya	69	7.4	16.2	15.6	8.9	5.2	39.7	81.6	25	9.9	3.07	23.5	13.6
38	An	82	41.2	37.9	21	14.4	14.7	37.5	86.1	28.9	12	3.48	25.2	7.1
39	Taunggyi	42	75.7	91.1	75.5	58.7	72.5	57.8	83.5	13.1	4.9	2.20	26.2	69.4
40	Pindaya	55	43.6	75.6	50.5	16.2	51.8	80.1	90.2	9.3	3.4	2.40	25.7	13.6
41	Ywarngan	60	23.7	78.7	26.1	12	47.1	71.8	90.2	9.2	4.3	2.26	25.4	5.5
42	Yatsauk	64	47.6	65.3	63	32.1	60.5	64.5	90	18.6	8.8	2.48	24.5	23.3
43	Loilin	52	54.8	59.3	55.3	25.3	73.2	66.8	87	40.4	20.7	2.87	23	25.7
44	Kunhing	85	36.5	72.8	47.3	34	70.8	62	88.8	48.9	30.3	2.76	28	34.7
45	Mineshu	73	33.7	35.2	46.5	23.1	72.4	76.9	92.9	56.6	41.7	3.03	25.4	23
46	Mone	70	57.6	58.1	51.6	26.8	74.5	60.7	87.1	47.2	34.8	3	25.8	34.1
47	Kukai	48	60.6	66.5	47.4	26.5	65	46.4	82	32.8	22.3	3.43	21.3	24.6
48	Naung Khio	55	55	70.1	54.8	32.3	69.4	72.6	92	11.4	5.7	2.48	26.4	12.6

APPENDIX (A) (cont.)

49	Namsan (North)	76	14.1	52.6	39.1	9.8	50.3	82.7	90	33.1	17.6	3.28	24.7	5.7
50	Manton	94	33.4	25.3	18.8	5.5	36.1	73.8	87.5	53.5	32.8	4.75	21.3	10.6
51	Laukine	12	49.8	52.7	77.9	71.7	73.8	48.6	85.7	59.6	43.5	3.29	22.6	26.1
52	Hopan	39	68.7	50	55.7	48.8	58.6	66.8	90.1	65.6	56.7	3.33	22.3	30.8
53	Minelar	30	73.3	68.2	67.3	55.5	85	77.1	88.5	70.3	47.2	2.50	26.8	47.5
54	Mineyaung	52	92.5	88.8	73.9	32.4	88.8	67.7	92.1	76.1	39.8	2.45	30.7	17.7

Source: 2014 Population and Housing Census, Department of Population.

APPENDIX (B)

SPSS Output

Principle Component Analysis

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.774
Bartlett's Test of Sphericity	Approx. Chi-Square	744.952
	df	66
	Sig.	.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.982	49.849	49.849	5.982	49.849	49.849	4.158	34.649	34.649
2	2.366	19.719	69.568	2.366	19.719	69.568	3.392	28.267	62.915
3	1.558	12.985	82.554	1.558	12.985	82.554	2.357	19.638	82.554
4	.640	5.333	87.886						
5	.564	4.700	92.587						
6	.266	2.218	94.804						
7	.243	2.025	96.830						
8	.208	1.731	98.560						
9	.081	.679	99.240						
10	.047	.394	99.633						
11	.026	.218	99.852						
12	.018	.148	100.000						

Extraction Method: Principal Component Analysis.

Rotated Component Matrix

	Component		
	1	2	3
PHCD	.948	.234	-.075
PHMF	.929	.247	.082
POUP	.913	.184	-.180
MLFPR	-.330	-.089	.562
POMI	.104	-.784	.243
POFI	.029	-.770	.258
TFR	-.496	-.756	-.002
MA	.378	.747	-.033
PHIS	.539	.705	.085
PHID	.561	.609	.207
PHMV	.020	-.161	.837
FLFPR	-.155	.060	.799

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser

Normalization.^a

a. Rotation converged in 5 iterations.

Regression

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	Urban society, Household's education status, Economic status		Enter

a. Dependent Variable: IMR

b. All requested variables entered.

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.851 ^a	.724	.708	12.22322	1.669

a. Predictors: (Constant), Urban society, Household's education status, Economic status

b. Dependent Variable: IMR

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	19627.127	3	6542.376	43.789	.000 ^b
	Residual	7470.354	50	149.407		
	Total	27097.481	53			

a. Dependent Variable: IMR

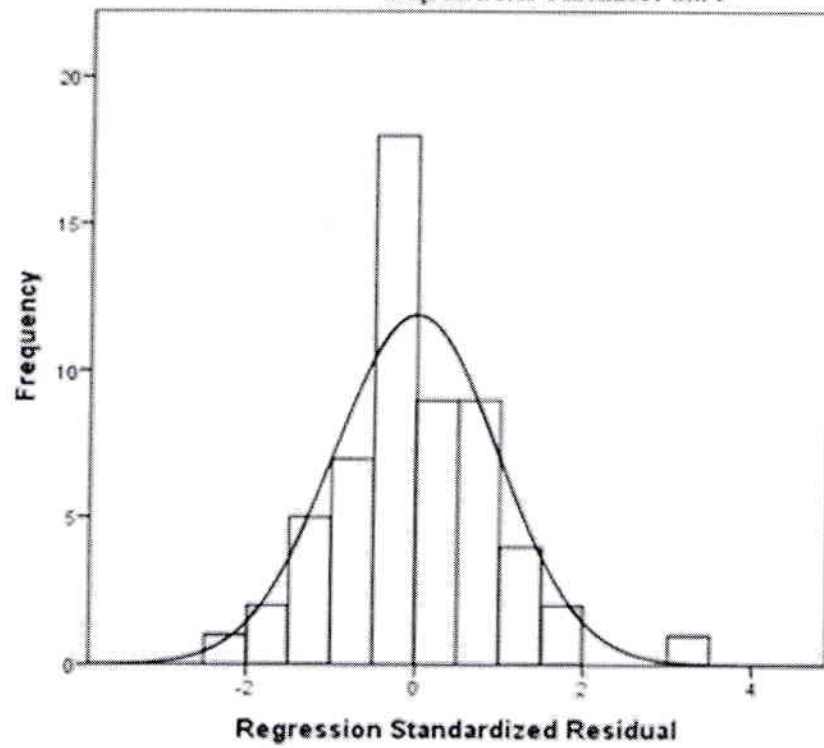
b. Predictors: (Constant), Urban society, Household's education status, Economic status

Coefficients

Model		Unstandardized Coefficients		t	Sig.	Collinearity Statistics	
		B	Std. Error			Tolerance	VIF
1	(Constant)	54.519	1.663	32.776	.000		
	Urban society	-18.302	1.679	-10.900	.000	1.000	1.000
	Household's education status	-5.499	1.679	-3.275	.002	1.000	1.000
	Economic status	-2.268	1.679	-1.351	.073	1.000	1.000

Histogram

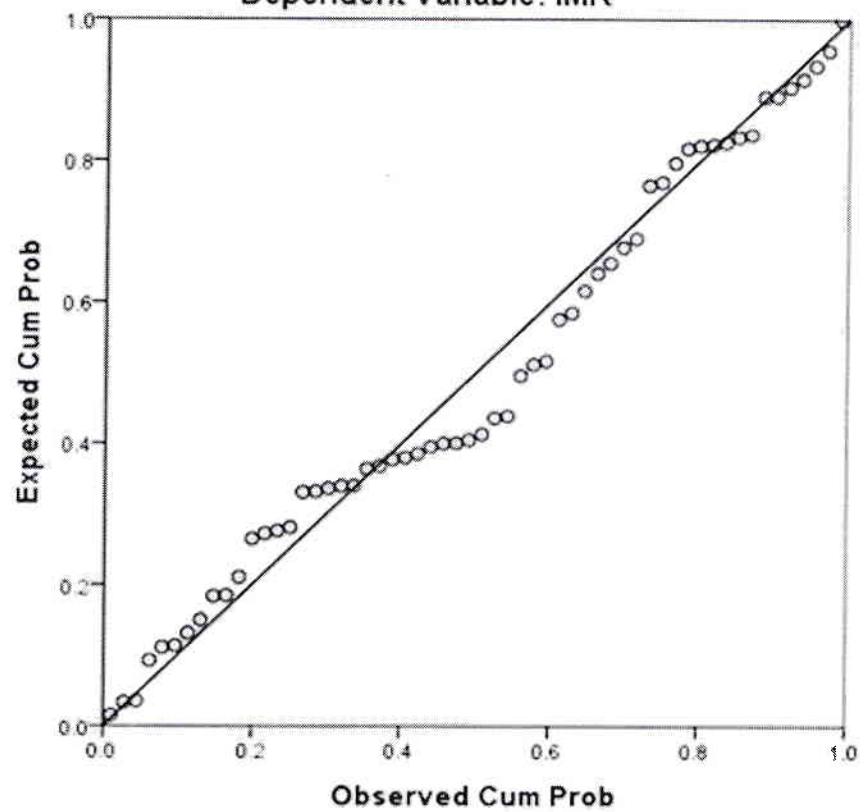
Dependent Variable: IMR



Mean=-1.39E-16
Std. Dev= 0.973
N=54

Normal P-P Plot of Regression Standardized Residual

Dependent Variable: IMR



Scatterplot

Dependent Variable: IMR

