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

L. U. KALPAGE

(158378A)

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

Department of Mathematics

University of Moratuwa

Sri Lanka

January 2020

Declaration of the Candidate and the Supervisor

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

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

Signature: Date:

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

Signature of the supervisor: Date:

Abstract

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

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

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

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

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

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

Acknowledgements

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

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

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

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

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

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

TABLE OF CONTENTS

Declaration of the Candidate and the Supervisor i
Abstract ii
Acknowledgements iii
Table of contents iv
List of figures vi
List of tables vi
List of abbreviations vii
List of appendices vii
Chapter 1
Introduction 1
1.1 Overview 1
1.2 Background of the Study 1
1.3 Statement of the Problem 2
1.4 Objectives of the Study 3
1.5 Significance of the Study 3
1.6 Organization of the Report 3
Chapter 2
Literature review 5
2.1 Overview 5
2.2 Poverty in Sri Lanka 5
2.3 Literature Review of Researches Related to Analysis of Poverty 7
2.3.1 Researches related to the poverty in Sri Lanka 7
2.3.2 Researches related to poverty in foreign countries 9
2.4 Summary 11
Chapter 3
Methodology 12
3.1 Overview 12
3.2 Source of Data 12
3.3 Data Description 12
3.4 Research Methodology 15
3.5 Brief Description of Statistical Techniques Used in the Study 16
3.5.1 Binary logistic regression 16
3.5.2 One Sample Proportion Test 22

3.5.3 Chi-Square Test 22
3.5.4 Mann-Whitney U Test 23
Chapter 4
Data analysis 24
4.1 Overview 24
4.2 Descriptive Analysis 24
4.2.1 Composition of the sample 24
4.3 Advanced Analysis 29
4.3.1 Reduction of factor levels 29
4.3.2 Identification of the association between independent variables and dependent variable 31
4.3.3 Detecting multicollinearity in binary logistic regression 33
4.3.4 Checking linearity assumption 34
4.3.5 Binary logistic regression model 36
Chapter 5
Conclusions and Recomendations 48
5.1 Overviews 48
5.2 Conclusions 48
5.3 Recommendation 50
REFERENCES 51
APPENDICES 51

LIST OF FIGURES

Figure 4.1: Classification of poverty status by province 26
Figure 4.2: Classification of poverty status by District 27
Figure 4.3: Log odds of poverty status Vs age of the household head 34
Figure 4.4: Log odds of poverty status Vs household size 35
Figure 4.5: Log odds of poverty status Vs income of the household 35
Figure 4.6: Log odds of poverty status Vs THHI 36
Figure 4.7: Plot of sensitivity and specificity Vs all possible cut point 43
Figure 4.8: ROC Curve 44
Figure 4.9: Plot of leverage Vs estimated logistic probability 46
Figure 4.10: Plot of delta beta Vs estimated logistic probability 46

LIST OF TABLES

Table 2.1: Information on poverty by country, residential sectors, provinces and districts in the year 2016 5
Table 3.1: Description of the variables 13
Table 4.1: Composition of the sample of HIES 24
Table 4.2: Composition of the sample taken for the study 24
Table 4.3: Classification of Poverty status by residential sector 25
Table 4.4: Classification of Poverty status by education level of the household head 28
Table 4.4: Test results of the one sample proportion test 29
Table 4.5: Description of new factor levels 30
Table 4.6: Identification of the association between independent variables and dependent variable 32
Table 4.7: Diagnostics tests results for selected models 33
Table 4.8: Baseline Model 37
Table 4.9: Details of developed model 38
Table 4.10: Description of the forward steps 41
Table 4.11: Omnibus Tests of Model Coefficients 42
Table 4.12: Model summary 42
Table 4.13: Classification table 44
Table 4.14: Details of the area under the ROC Curve 45
Table 4.15: Hosmer and Lemeshow test 45
Table 4.16: Classification table for model validation 47

LIST OF ABBREVIATIONS

CI - Confidence Interval

CPH - Census of Population and Housing

DCS - Department of Census and Statistics

HIES - Household Income and Expenditure Survey

ML - Maximum likelihood

OPL - Official Poverty Line

OR - Odds Ratio

ROC - Receiver Operating Characteristic

SDG - Sustainable Development Goals

VIF - Variance Inflation Factor

LIST OF APPENDICES

Appendix	Description	Page
Appendix - A	Cross tabulation on poverty status by Province	54
Appendix - B	Cross tabulation on poverty status by District	56
Appendix - C	Normality test for continuous variables with Poverty Status	59
Appendix – D	Logistic regression model with 13 variables including Household Income	60