

THE MAJOR CAUSES OF THE RURAL POVERTY

A CASE STUDY OF KENYA¹⁾

BY

Martin M. Kasina

SCHOOL OF MATHEMATICS
COLLEGE OF BIOLOGICAL AND PHYSICAL
SCIENCES
UNIVERSITY OF NAIROBI

A PROJECT SUBMITTED TO THE SCHOOL OF MATHEMATICS IN PARTIAL
FULFILLMENT FOR THE DEGREE OF MASTER OF SCIENCE IN SOCIAL
STATISTICS.

©2009

University of NAIROBI Library



0378987 2

Declaration

I the undersigned, declare that this project proposal is my original work and to the best of my knowledge has not been presented for the award of a degree in any other University


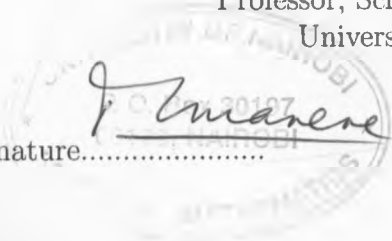
Martin M. Kasina
Reg. No. 156/70849/07

Signature.....


Date.....12/8/2009

This project proposal has been submitted for registration with our approval as University Supervisors

Prof. Manene
Professor, School of Mathematics
University of Nairobi

Signature.....


Date.....13/8/2009

Dr. Weke
Lecturer, School of Mathematics
University of Nairobi

Signature.....


Date.....13/08/2009

Aknowledgement

I wish to greatly acknowledge and appreciate my Supervisors, Prof.Manene and Dr.Weke for their scholarly guidance and commends in writing this paper.Thanks to all my lecturers;Dr Achia, Dr Nguti, Prof.JAM, Prof Owino, Anne Wang'ombe, Obudho and Odwesso and coursemates particularly Francis L(Tanzania), Francois N. (Rwanda),Mwaniki A. and George W. for their continued surpport and encouragement.

Dedication

Dedicated to

- My wife Loise and my son Charis
- My parents.

Abstract

Poverty is a multi-dimensional phenomenon whose causes are complex and difficult to Measure precisely. The need to articulate causes of poverty in developing countries is Paramount given its effects. This study examines the major causes of rural poverty from 942 households. For poverty reduction strategies to be effective, they must be closely linked with the objectives of the poverty alleviation program. Most of the official poverty statistics released including the data statistics used in this study focuses on the poverty incidence, the income gap, and the poverty gap and traps. However, identifying the characteristics of vulnerable families and the vulnerable provinces is important towards understanding the causes of their vulnerability and in formulating programs on poverty reduction. This study proposed a framework and a methodology for identifying the sub locations, districts and the provinces where these families are located and the indicators of a potential deprived community to guide program implementers in the allocation of resources. In this study logistic regression analysis was applied. The results indicated that family size, geographical location, agriculture knowledge, empowerment and the gender of the head of the household as the key determinant of the probability of one being poor in a rural set up. Age is factor when run as a univariate logistic regression but when run in a multivariate logistic regrassion it is not a predictor of ones probability of being poor.

List of Figures

4.1	Sex distribution	26
4.2	Age group Distribution	27
4.3	Age group	28
4.4	Households Distribution in the Provinces	29
4.5	Households Distribution in the Districts	30
4.6	Agriculture Knowledge source	31
4.7	Empowered (Rights)	32
4.8	Poor Vs Better off	33
4.9	Dailly household income	36

List of Tables

3.1	Variables used in the Data	24
4.1	Sex Distribution	26
4.2	Age group Distribution	27
4.3	Age spread	27
4.4	Households distribution in the Provinces	29
4.5	Households Distribution in the Districts	30
4.6	Agriculture Knowledge Source	31
4.7	Empowered(Rights)	32
4.8	Poor Vs Better off	33
4.9	Crosstabulations	34
4.10	Average income per day per household	35
4.11	Crosstabulations	37
4.12	Crosstabulations	37
4.13	Crosstabulations	37
4.14	logistic analysis1	39
4.15	logistic analysis2	39
4.16	logistic analysis3	40
4.17	logistic analysis4	41
4.18	logistic analysis5	41
4.19	logistic analysis6	41

Contents

1	INTRODUCTION	2
1.1	Background to study	2
1.1.1	poverty measure	3
1.2	statement to the problem	3
1.3	Objectives	4
1.3.1	General objective to the study	4
1.3.2	specific objective of the study	5
1.4	Statement of Hypothesis	5
1.5	Research questions	6
1.6	Significance of the study	6
1.7	Limitations of the study	7
1.8	Delimitations of the study	7
1.9	Defination of terms used in the study	7
1.10	Organization of the Study	8
2	LITERATURE REVIEW	9
3	METHODOLOGY	12
3.1	The Data	12
3.2	Logistic Regression Models	12
3.2.1	Goodness of fit	17
3.2.2	Model selection	20
3.2.3	Assumption of the logistic Regression model	20
3.3	FGT (<i>Foster, Greer and Thobeck Method</i>)	21
3.4	Data Analysis	23

4	DATA ANALYSIS, PRESENTATION AND INTERPRETATIONS	25
4.1	Biodata of the respondents	26
4.1.1	sex	26
4.1.2	Age	27
4.2	Households Distribution	29
4.3	Crosstabulations	33
4.4	Odds Ratio	39
4.4.1	Univariate Logistic Regression	39
4.4.2	Multivariate logistic regression	42
5	SUMMARY OF FINDINGS, CONCLUSION AND RECOMMENDATIONS	45
5.1	Summary of the findings	45
5.2	conclusion	47
5.3	Recommendations	48

Chapter 1

INTRODUCTION

1.1 Background to study

Poverty is a state of deprivation, measured at various levels of society: individual, household, community, national, and international; and that a view of this deprivation dependent on one's discipline. Poverty is multi-dimensional and that policies aimed at fighting poverty require credible evidence on poverty traps and their causes. Poverty, being a multidimensional issue required a mixed approach in its measurement before we reduce indicators of well-being into money metrics or a single index. The multidimensional nature of poverty requires that we move away from one index since we can look at all aspects of deprivation and be able to tell who is poor or not. Some dimensions have been neglected in the past studies like empowerment, access to market at village level and access to formal agricultural knowledge and thus why this study is very crucial and timely.

1.1.1 poverty measure

Poverty can be measured differently and each emphasizes a different aspect of poverty.

examples

1. **Poverty line** :This approach emphasizes income, mostly used at national level of poverty measure
2. **Minimum Basic needs**:This approach looks at poverty as a form of deprivation of survival, security and enabling needs eg literacy level.
3. **Human Development Index (HDI)**:This approach Focuses on life expectancy, Knowledge and real per capita income, mostly used at global level of poverty measure and comparison
4. **Land based poverty** :This approach looks at poverty as poor infrastructure and insecurity of tenure.

1.2 statement to the problem

In the past two decades, Kenya has been experiencing an economic decline. The situation has continued to deteriorate and the country is now among the world's poorest countries (World Bank: 1993, 2000). Although most of the studies done in this area concur in their findings that most vulnerable groups include landless, pastoralists and female-headed households in the country (Kenya), none of them has paid critical attention to factors that contribute to this state of affairs. Admittedly, their consideration of such factors is casual at best. Most of the

findings addresses the symptoms rather than causes of continued poverty in the country. This study therefore fills this gap by investigating economic and social factors responsible for the persistently high poverty levels among the households in Kenya despite many programs and campaigns to fight poverty and inequality.

1.3 Objectives

1.3.1 General objective to the study

The purpose of this study is to determine the major causes of rural poverty in Kenya, through the analysis of the data collected by FASID, ICRAF and TENGEMEO INSTITUTE in 2004 by using logistic regression.

1.3.2 specific objective of the study

- To ascertain, establish and predict the demographic characteristics of the rural poor.
- To identify if there is any statistically significant relationship between the predictors(Independed Variable) and poverty level(Depended Variable).

1.4 Statement of Hypothesis

$$H_0 : \beta_{xi} = 0$$

$$H_a : \beta_{xi} \neq 0 \text{ Where}$$

X_1 =Gender of the household head

X_2 =Family size

X_3 =Age of household head

X_4 =Location of household(District)

X_5 =Empowerment(right to decide alone)

X_6 =Agriculture Knowledge source

And β is the change of dependent variable(DV) due to a unit change of independent(IV) ceteris puribus The researcher shall do logistic regression to find out how change of an independent variable by a unit affects poverty, holding all other variables constant. Much of this will be discussed in chapter four. The total income per household per day will be determined. If a household earns less than a dollar per day i.e Ksh.65.55 (the exchange rate in 2004) it will be classified as poor otherwise it

is better off.

1.5 Research questions

- Does farmers' agricultural knowledge source affect their agricultural technology adaptation?
- Does ones location predetermine S.E.S (*Social economic status*)?
- Is there a relationship between the level of poverty and the gender of household head?
- Does the household head level of education influence adaptation of agricultural technology and poverty level?.

1.6 Significance of the study

Developmental interventions are unlikely to achieve their desired outcomes, as they are implemented as intended (Basch, 1984) and (Botin, 1990). It should not be assumed that once a country innovates, full implementation would follow (Basch, 1984) (Resnicow, Cross, and Wynder, 1993). In general, interventions are more likely to be implemented successfully if they are relatively simple to understand and deliver, do not require a lot of resources and the need is well diagnosed (Rogers, 1983). Therefore thorough diagnosis of the obstacles to poverty eradication is core to any successful program implementation for poverty reduction in our country. Hence, this study will be important because it will explore on the adequacy of the farmers' knowledge at village level and discover barriers that hinder effective implementation of most of the poverty eradication programs.

1.7 Limitations of the study

The poverty index used in this study will be arrived at by using the household income and property owned but some of the income is from the business which may not be consistent. Therefore Due to this some of the observations might not be the true reflection of the household' social economic level. Again some institutions important to traditional societies have lingered, inhibiting the transition to new techniques of production. The resistance of norms and institutions to change is one reason why the outcome might be disappointing.

1.8 Delimitations of the study

The data used in this study is limited to 942 households only, majority in a rural set up, although the study is addressing the whole country. The findings apply to the visited households. The households were randomly selected.

1.9 Defination of terms used in the study

- Vulnerability has been referred to as ex-ante poverty
- Poverty incidence is a ratio of two counts: total number of poor households and total number of households.
- A Poverty Trap is a scenario 'in which a poor country is simply too poor to achieve sustained economic growth, The trap becomes cyclical and begins to reinforce itself if steps

is better off.

1.5 Research questions

- Does farmers' agricultural knowledge source affect their agricultural technology adaptation?
- Does one's location predetermine S.E.S (*Social economic status*)?
- Is there a relationship between the level of poverty and the gender of household head?
- Does the household head level of education influence adaptation of agricultural technology and poverty level?.

1.6 Significance of the study

Developmental interventions are unlikely to achieve their desired aims unless they are implemented as intended (Basch, 1984) and (Botvin, 1990). It should not be assumed that once a country adopts an innovation, full implementation would follow (Basch, 1984) and (Resnicow, Cross, and Wynder, 1993). In general, interventions are more likely to be implemented successfully if they are relatively simple to understand and deliver, do not require high levels of resources and the need is well diagnosed (Rogers, 1995). Therefore thorough diagnosis of the obstacles to poverty alleviation is core to any successful program implementation for poverty reduction in our country. Hence, this study will be important because it will explore on the adequacy of the farmers' inputs at village level and discover barriers that hinder effective implementation of most of the poverty eradication programs.

1.7 Limitations of the study

The poverty index used in this study will be arrived at by using the household income and property owned but some of the income is from the business which may not be consistent. Therefore Due to this some of the observations might not be the true reflection of the household' social economic level. Again some institutions important to traditional societies have lingered, inhibiting the transition to new techniques of production. The resistance of norms and institutions to change is one reason why the outcome might be disappointing.

1.8 Delimitations of the study

The data used in this study is limited to 942 households only, majority in a rural set up, although the study is addressing the whole country. The findings apply to the visited households. The households were randomly selected.

1.9 Defination of terms used in the study

- Vulnerability has been referred to as ex-ante poverty
- Poverty incidence is a ratio of two counts: total number of poor households and total number of households.
- A Poverty Trap is a scenario 'in which a poor country is simply too poor to achieve sustained economic growth, The trap becomes cyclical and begins to reinforce itself if steps

are not taken to break the cycle ie vicious circle of poverty is developed

- Income gap/disparity or wage gap is a term used to describe inequities and asymmetry in the distribution of wealth and income between socio-economic groups within society
- Headcount refers to the proportional population of the poor to the total populaation
- Agricultural technology refers to application of fertilizers, breded cows and artificial animal feed

1.10 Organization of the Study

The study will be organized into five chapters:

1. The first chapter will deal with the introduction to the study which has the background to the study, statement of the problem, purpose of the study, research objectives, research questions, limitations of the study, delimitations, basic assumptions of the study, definition of significant terms and lastly organization of the study.
2. Chapter two will present the review of related literature.
3. In Chapter Three, methodology of the study will be presented.
4. Chapter Four will consists of data analysis, research findings, presentation and discussion of the findings.
5. In Chapter Five, the summary of the findings, conclusions and recommendations for further research will be presented.

Chapter 2

LITERATURE REVIEW

Ordinarily, one would wish not to discuss such a depressing subject as poverty. However, like death and taxes, it has become a continuing affliction with which many Kenyans and Africans in general have to contend with on a daily basis (Ayako, et al, 1997). It is argued that a sixth of the world's people produce 78 percent of its goods and services, receiving about 78 percent of the world's income. Three-fifths of the world's population live in the poorest 61 countries where they receive only 6 percent of the world's income (World Bank, 2000). This scenario could be an understatement as poverty goes beyond income measurements. Besides low income poverty manifests itself in the form of malnutrition; high mortality rate; illiteracy; lack of access to basic education, drinking water, main health facilities and shelter. It is a problem that manifests itself in many forms and at all levels of the society threatening the very foundations of the society. The main forms of poverty include deprivation, isolation, alienation, insecurity and worst of all, despondency. Poverty is defined as lack of income required for households to meet their basic needs. This emerges from the utilitarian and welfare perspectives that explain poverty purely in terms of income-expenditure

and nutrition-consumption deficiencies. Contemporary thought asserts that poverty is caused by a diverse set of power relationships that deny life skills, assets and resources to people. Access and control over productive resources both private assets and public services form the basis of all entitlements. (Paul Kamau, University of Nairobi ,2006). Studies have found that female-headed households constitute 30 percent of Kenya's households of whom 52.9 percent are poor (Mukui, 1994; Republic of Kenya, 1999; World Bank, 1980). In spite of comprising a clean majority of the population, women have lived a marginalized life in all spheres. The hard work by these women who head households notwithstanding, their incomes are limited and not always certain (Republic of Kenya, 1995). Circumstances compel many of them to engage in illegal and usually risky activities such as prostitution, brewing of illicit liquor, hawking without license and in restricted places. In effect, their children are potentially at great risk of neglect and abuse upon these women's arrest and imprisonment. Ultimately, these children find their way into the streets thus forming part of street children. Ayako, et al (1997) observes that poverty in Kenya is pervasive and widespread among socio-economic groups. The major socio-economic groups amid which the poor are found are the female-headed households (53 percent), subsistence farmers (47 percent), food crop farmers (46 percent), pastoralists (36 percent), and private sector workers (31 percent). They argue that dualism, population pressure, state superstructure and policy bias, poor natural resource base, gender bias, natural disasters, exploitative intermediaries and international processes cause poverty. It has been established that some laws hinder the development of single women in several areas. Such laws that clearly constrain women participation

in development especially in the small-scale enterprises include the Employment Act and the Succession Act. Admittedly, a review of these Acts will assist women inherit land and property, the most commonly used form of security for borrowing. Based on the general literature and the country's Participatory Poverty Assessment (PPA) study reports, poverty in Kenya appears to be an economic, sociological, psychological and political phenomenon, which manifests itself in vulnerability, insecurity, isolation, alienation, domination and dependence, material deprivation, denial of freedom of choice, and lack of participation and assets.

The most recent study on this topic was done by Nancy Khadioli 2008, School of mathematics, University of Nairobi. She found the demographic factors that increase the probability of being poor are: the household head age, religion, region and ethnicity. However she did not look at the forces underlying those outcomes. She used the asset index to measure the economic welfare of the households but many changes may take place in the economic situation of many households and the asset indices remain the same. Again you find that most of the assets included in the asset index are by their nature urban rather than rural. Therefore the ideal method to measure the household economic welfare is income and assets indices to arrive at the poverty index.

Chapter 3

METHODOLOGY

3.1 The Data

The data used to analyze the major causes of rural poverty is taken from 2004 ICRAF, FACID AND TENGEMEO INSTITUTE for Kenya. It is nationally representative covering 7124 women and 3176 men selected from 942 sample points (clusters) throughout Kenya. The survey collected information related to demographic and detailed information on livestock ownership, insecurity, accessibility to market, gender and educational level of the household, household expenditure and accessibility to credit facilities. It utilized four stage sample design. That is selection of the province, District, the village and the sample points (clusters) from a national master sample maintained by the kenya national bureau of statistics (KNBS) throughout Kenya.

3.2 Logistic Regression Models

The present study used the data from the Fasid, Icraft and Tenge-meo institute of 2004 from 942 households, research on poverty,

environment and agricultural technology. Logistic regression was used test to the the major causes of the rural poverty. The dependent variable being the dichotomous(binary) variable of whether the household is poor(1) or not (0) that is the social economic status; logistic regressions allow us to have more than one predictor in our model. We shall also estimate the association between each predictor and $\Pr(y = 1)$ controlling for all other predictors

$$y = \begin{cases} 1, (\text{poor}) \\ 0, (\text{not poor}) \end{cases}$$

Where y is the SES as either poor or better off.

The central mathematical concept that underlies logistic regression is the logit-the natural logarithm of an odds ratio. Logistic regression describes the relationship between a dichotomous response variable and a set of explanatory variables. The explanatory variables may be continuous or discrete(with dummy variables).

We take y_i as random variable that can take the values 1 and 0 with probabilities π and $1 - \pi$ respectively such that,

$$pr[y_i = 1 | x_1, x_2, \dots, x_k] = \pi \tag{3.1}$$

$$pr[y_i = 0 | x_1, x_2, \dots, x_k] = 1 - \pi \tag{3.2}$$

Where x 's are the predictor variables in our model. The simple logistic model has the form

$$\text{logit}(\pi) = \text{natural logs(odds)} = \ln \frac{\pi}{1-\pi} = \alpha + \beta X \text{ Therefore our}$$

link function is $\ln \frac{\Pi}{1-\Pi}$ such that

$$\ln \frac{\Pi}{1-\Pi} = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k \quad (3.3)$$

Whereby β denotes a vector of parameters, X is often called the design matrix and $X\beta$ is the linear component of the model, Such that,

$$\beta X = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (3.4)$$

$$\Pi = (y = 1 | X = x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \quad (3.5)$$

Where y is the outcome of Interest.

Note that the odds vary on a scale of $[0, \infty]$, so that the log odds can vary on the scale of $[-\infty, \infty]$. To ensure that π is restricted to the interval $[0, 1]$ it is often modelled using a cumulative probability distribution

$$\pi = \int_{-\infty}^t f(s) ds$$

Where $f(s) \geq 0$ and $\int_{-\infty}^{\infty} f(s) ds = 1$

The probability density function $f(s)$ is called the *tolerance distribution*

Exponentiating equation (3.3) we get

$$\exp(\text{logit}(\pi)) = \exp(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k) \quad (3.6)$$

$$\frac{\pi}{1 - \pi} = \exp(\beta_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k) \quad (3.7)$$

$$\Pi = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \quad (3.8)$$

Hence our logistic or logit model is

$$\pi = \text{In} \frac{\pi}{1 - \pi} = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \quad (3.9)$$

$$\text{logit} \pi_i = \text{In} \frac{\pi_i}{1 - \pi_i} = X_i^T \beta$$

Therefore our link or transformation function is $\text{In} \frac{\pi}{1 - \pi}$ whereby $\text{logit}(\pi) = \text{In} \frac{\pi}{1 - \pi}$ is the logistic transformation of the probability of poor. Therefore by letting

$$\sum_{j=0}^k \beta_j X_j = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k = Z \quad (3.10)$$

Then we can now have the logistic function which is the inverse of the logit function obtained as

$$\frac{\pi}{1 - \pi} = e^z \Rightarrow \pi = (1 - \pi)e^z \Rightarrow \pi = \frac{e^z}{1 + e^z} \quad (3.11)$$

Note that our model has unknown parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ which must be estimated in order to fit the logistic regression model to our data. Maximum likelihood will be used to estimate these parameters

The likelihood function is given by

$$L = \prod_{i=1}^n \binom{n_i}{y_i} \pi_i^{y_i} (1 - \pi_i)^{n_i - y_i} \quad (3.12)$$

The log-likelihood function is given by

$$\begin{aligned} \ln L &= \sum_{i=1}^n \left[\ln \binom{n_i}{y_i} + y_i \ln \pi_i + (n_i - y_i) \ln (1 - \pi_i) \right] \quad (3.13) \\ &= \sum_{i=1}^n \left[\ln \binom{n_i}{y_i} + y_i \ln \frac{\pi}{1 - \pi_i} + n_i \ln (1 - \pi_i) \right] \\ &= \sum_{i=1}^n \left[\ln \binom{n_i}{y_i} + y_i z - n_i \ln (1 + e^z) \right] \end{aligned}$$

Where $Z_i = \sum_{j=0}^k \beta_j x_{ij}$. The derivative of the log-likelihood function w.r.t the unknown β parameters are

$$\frac{\delta \ln L}{\delta \beta_j} = \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n n_i x_{ij} \frac{e^z}{1 + e^z}, j = 0, 1, \dots, k \quad (3.14)$$

Therefore our betas will be estimated as, $b = (X^T X)^{-1} (X^T y)$

Where X is the matrix of our predictors and y is the poverty index.

3.2.1 Goodness of fit

We can use the likelihood ratio test whereby the ratio of the maximized value of the likelihood function for the full model \hat{L}_f over the maximized value of the likelihood function for the simpler model or current model \hat{l}_c . The likelihood-ratio test statistic equals:

$$-2 \log \left(\frac{L_0}{L_1} \right) = -2[\log(L_0) - \log(L_1)] = -2(L_0 - L_1) \quad (3.15)$$

By comparing \hat{l}_c and \hat{l}_f , the extent to which the current model adequately represents the data can be judged. A comparison between \hat{l}_c and \hat{l}_f can be made on the basis of the deviance statistic which is given by

$$D = -2In\frac{\hat{l}_c}{\hat{l}_f} = -2[In\hat{l}_c - In\hat{l}_f] \quad (3.16)$$

Deviance statistic involves a comparison of the log of the maximum of the saturated model and the assumed model. when \hat{l}_c is small relative to \hat{l}_f , then the value of D is large, indicating that the current model is poor. small value of D are encountered when \hat{l}_c is similar to \hat{l}_f indicating that the current model is a good one. In modeling n binomial observations, the likelihood function is

$$L = \prod_{i=1}^n \binom{n_i}{y_i} (1 - \pi_i)^{n_i - y_i} \pi_i^{y_i}$$

on fitting a logistic model with $k + 1$ Unknown parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_k$, fitted values π_i are obtained where

$$\text{logit}(\pi) = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \dots + \hat{\beta}_k X_{ki}$$

The maximum log-likelihood under this model is given by

$$\text{In}\hat{L}_c = \sum_{i=1}^n [\text{In}\binom{n_i}{y_i} + y_i \text{In}\hat{\pi}_i + (n_i - y_i) \text{In}(1 - \hat{\pi}_i)]$$

and so the maximum log-likelihood function for the full model is given by

$$\text{In}\hat{L}_f = \sum_{i=1}^n [\text{In}\binom{n_i}{y_i} + y_i \text{In}\pi_i^* + (n_i - y_i) \text{In}(1 - \pi_i^*)]$$

The deviance is then given by

$$\begin{aligned} D &= -2[\text{In}\hat{L}_c - \text{In}\hat{L}_f] \\ &= 2 \sum_{i=1}^n [y_i \text{In}\frac{\pi_i^*}{\hat{\pi}_i} + (n_i - y_i) \text{In}(\frac{1 - \pi_i^*}{1 - \hat{\pi}_i})] \end{aligned} \quad (3.17)$$

3.2.2 Model selection

Model selection is an important task in data analysis. The forward selection, backward elimination or stepwise logistic regression methods determines automatically which variables to add or drop from the model, R we preferred in our model selection. The binomial logistic in R offers; forward logistic regression, forward conditional, forward wald, backward conditional, backward logistic regression and backward wald. In our analysis we used stepwise logistic regression.

3.2.3 Assumption of the logistic Regression model

- logistic regression does not assume a linear relationship between the dependent and independent variable
- The dependent variable need not to be normally distributed, but does assume its distribution is within the range of the exponential family of distribution.
- The dependent variable need not to be homoscedastic for each level of the independent variable (no homogeneity of vari-

ance assumption)

- No assumption of normally distributed error terms.

3.3 FGT (*Foster, Greer and Thobek Method*)

This is a powerful computational tool for illustrating Head-count index(H), Poverty gap(PG) and Squared poverty gap index(SPG). Its major advantage is that it is sensitive to the distribution of the poor and severity of poverty.

$$FGT = P_{\alpha} = \int_0^z \left[\frac{z-x}{z} \right]^{\alpha} f(x) dx, \alpha \geq 0 \quad (3.18)$$

Where

x is Household consumption expenditure

$f(x)$ is Roughly the proportion of the population consuming x (its density)

z is poverty line

α is a Non negative parameter

Such that when;

- $\alpha = 0$ We get Headcount index (H)
- $\alpha = 1$ We get Poverty gap index(PG) and
- $\alpha = 2$ We get Squared poverty gap index(SPG) respectively

Equation 3.4 can also be written as

$$FGT = \sum_i^q \frac{1}{n} \left(\frac{z - y_i}{z} \right)^\alpha \quad (3.19)$$

FGT satisfies,

- Monotonicity axiom which states that decrease in poor person's income increases poverty index ceteris paribus
- Transfer axiom which states that pure transfer from a poor person to somebody with more income increases the poverty index ceteris paribus

3.4 Data Analysis

SPSS was used to keep and manage the data but R was used in data analysis and exploration for reporting.

No. of observations = 894		
	Variable	Class
1	Household ID	integer
2	Old Hh ID	numeric
3	Old count ID	factor
4	province	factor
5	district	factor
6	Hhsize	integer
7	Respondent ID	integer
8	Relation to HhH	factor
9	sex	factor
10	age	integer
11	business	factor
12	monthhome	numeric
13	whyhome	factor
14	bus income	numeric
15	prod income	integer
16	liv income	integer
17	milk income	numeric
18	income total	numeric
19	rights(Empowerment)	factor
20	Agriculture.Knowledge	factor
21	percapita	numeric
22	poverty index	factor
23	hsize1	integer
24	agegroup	factor

Table 3.1: Variables used in the Data

Chapter 4

DATA ANALYSIS, PRESENTATION AND INTERPRETATIONS

The following are the results of the data analysis and their interpretations. The analysis was done using R programme. 942 Households were interviewed however only 894 questionnaires were completely responded to (after cleaning the data) and therefore the findings applies to only these households. Contrary to our expectations, we did not observe any probability of being poor due age difference.

4.1 Biodata of the respondents

4.1.1 sex

Sex	Frequency	Percentage
Female	197	22
Male	697	78
Total	894	100

Table 4.1: Sex Distribution



Figure 4.1: Sex distribution

The above table shows the gender of the head of the household. Most of the heads are males

4.1.2 Age

age group	Frequency	Percent	Cum. percent
less29	12	1.3	1.3
30-44	176	19.7	21.0
45-54	224	25.1	46.1
55+	482	53.9	100.0
Total	894	100.0	100.0

Table 4.2: Age group Distribution

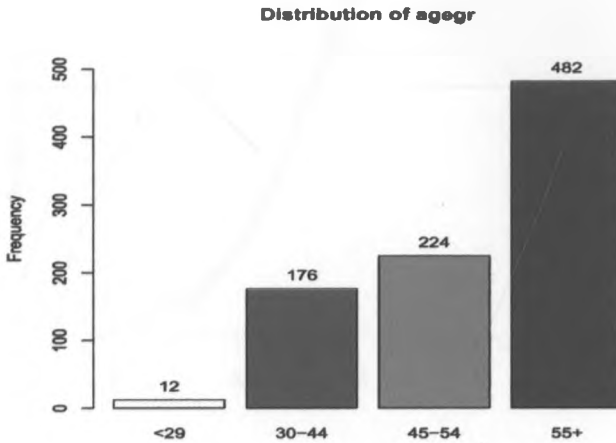


Figure 4.2: Age group Distribution

obs.	mean	median	s.d.	min.	max.
894	55.971	55	13.69	19	104

Table 4.3: Age spread

Distribution of age

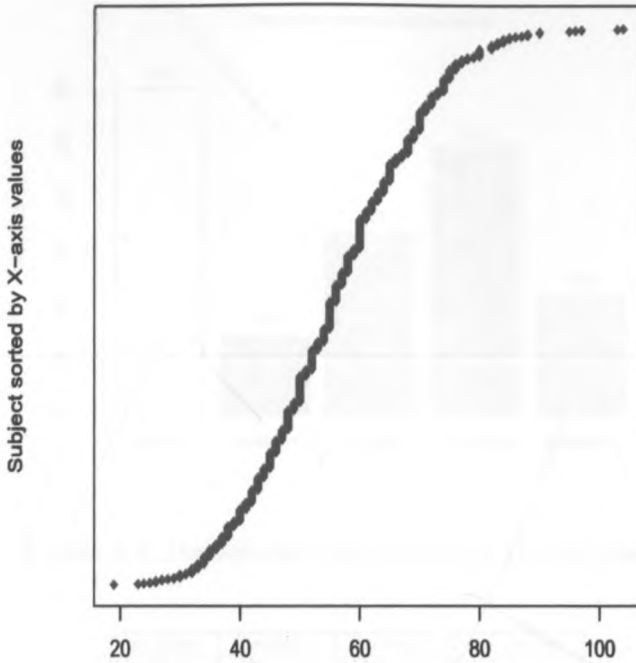


Figure 4.3: Age group

Households' heads were increasing with age majority being the group aged 50 and above years at 54%. This could be qualified by the fact that most families in rural areas are extended and the oldest is the household head most of the time. The maximum age was 104, median age 55 and mean age 55.97 years.

4.2 Households Distribution

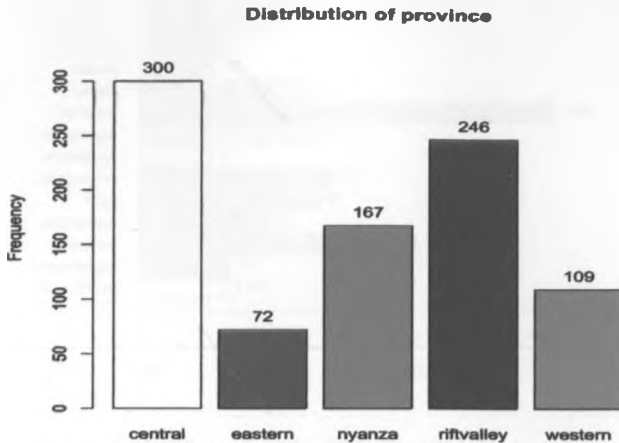


Figure 4.4: Households Distribution in the Provinces

province	Frequency	Percent	Cum. percent
central	300	33.6	33.6
eastern	72	8.1	41.6
nyanza	167	18.7	60.3
riftvalley	246	27.5	87.8
western	109	12.2	100.0
Total	894	100.0	100.0

Table 4.4: Households distribution in the Provinces

Central had the highest number of the households while Eastern had the fewest. The households were proportional to the population per the province. The figure below summarises the households distribution per district in our study.

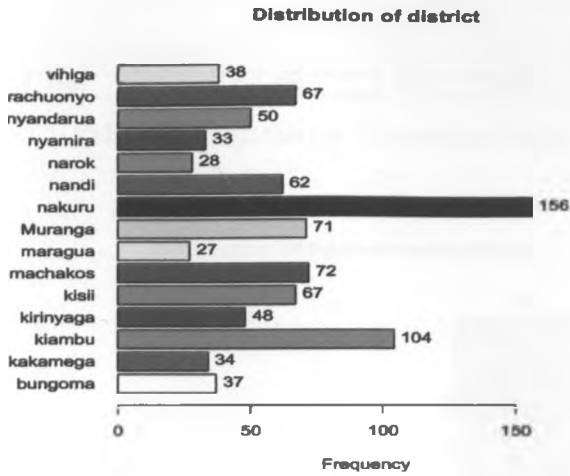


Figure 4.5: Households Distribution in the Districts

District	Frequency	Percent	Cum. percent
bungoma	37	4.1	4.1
kakamega	34	3.8	7.9
kiambu	104	11.6	19.6
kirinyaga	48	5.4	24.9
kisii	67	7.5	32.4
machakos	72	8.1	40.5
maragua	27	3.0	43.5
Muranga	71	7.9	51.5
nakuru	156	17.4	68.9
nandi	62	6.9	75.8
narok	28	3.1	79.0
nyamira	33	3.7	82.7
nyandarua	50	5.6	88.3
rachuonyo	67	7.5	95.7
vihiga	38	4.3	100.0
Total	894	100.0	10

Table 4.5: Households Distribution in the Districts

Nakuru had the highest number of households 156 while Kakamega had the least 34

Agriculture.Knowledge	Frequency	Percent	Cum. percent
official	136	15.2	15.2
others	145	16.2	31.4
parents	613	68.6	100.0
Total	894	100.0	100.0

Table 4.6: Agriculture Knowledge Source

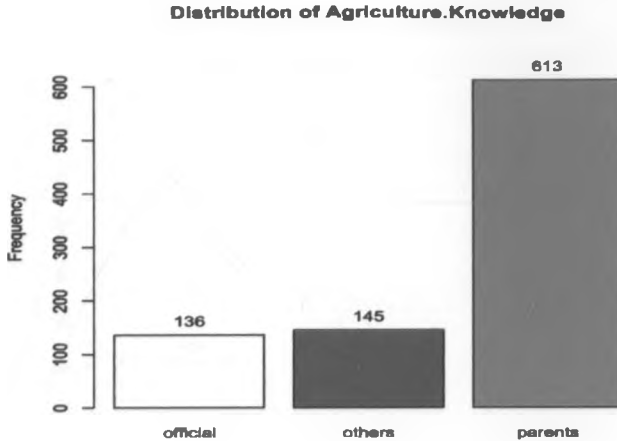


Figure 4.6: Agriculture Knowledge source

Most of the respondents receive the agricultural knowledge from the parents 613(69%), only 136(15%) received the agricultural knowledge. This is not very health since the parents may not be authoritative in that area.

rights	Frequency	Percent	Cum. percent
haverights	109	12.2	12.2
no rights	81	9.1	21.3
withapproval	704	78.7	100.0
Total	894	100.0	100.0

Table 4.7: Empowered(Rights)

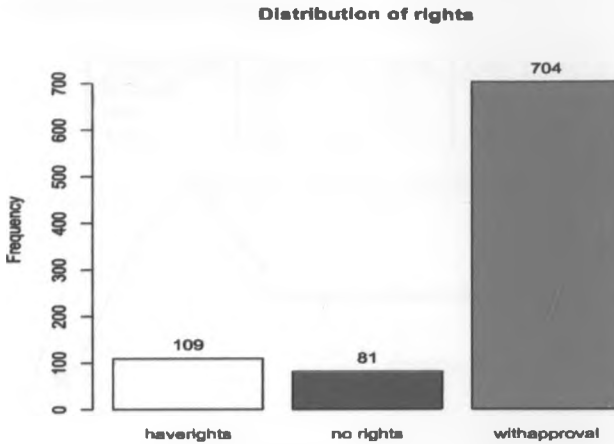


Figure 4.7: Empowered (Rights)

Those who have rights are 109(12%), the rest have no rights or they have to receive some approval before taking any action 785(88%)

4.3 Crosstabulations

Out of the 894 households interviewed 520(58%) were poor and 374(42%) were better off. The following tables and figures gives us the summary of our findings.

poverty index	Frequency	Percent	Cum. percent
betteroff	374	41.8	41.8
poor	520	58.2	100.0
Total	894	100.0	100.0

Table 4.8: Poor Vs Better off

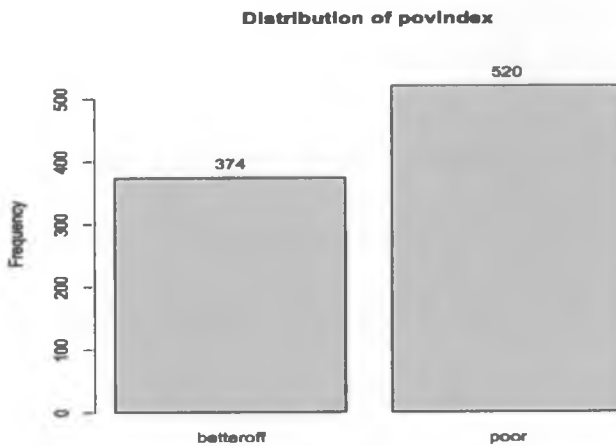


Figure 4.8: Poor Vs Better off

The following is a summary of crosstabulation for various variables and poverty index. The above table shows 76 females

Variable	betteroff	poor	Test stat.	P value
Sex			Chisq. (1 df) = 0.94	0.333
Female	76 (20.3)	121 (23.3)		
Male	298 (79.7)	399 (76.7)		
Age			Ranksum test	0.559
Median(IQR)	55 (46,65)	55 (46,67)		
Houseize			Chisq. (8 df) = 74.63	0.001
1	23 (6.1)	5 (1)		
2	31 (8.3)	6 (1.2)		
3	38 (10.2)	33 (6.3)		
4	45 (12)	50 (9.6)		
5	53 (14.2)	75 (14.4)		
6	55 (14.7)	69 (13.3)		
7	37 (9.9)	74 (14.2)		
8	39 (10.4)	59 (11.3)		
9 Above	53 (14.2)	149 (28.7)		

Table 4.9: Crosstabulations

were better off while 121 were poor. 298 males were better off while 399 were poor. The result is not statistically significant at 95% because the p-value 0.333 which is ≥ 0.05 . Again in the households with 1 member 23 households are better off while 5 are poor and so forth , the households with 9 and above memmmbers only 53 are better off with 149 poor. So it can be seen that poverty is increasing with the increase of members in the household.

obs.	mean	median	s.d.	min.	max.
894	93.432	54.666	130.51	0.625	1274.306

Table 4.10: Average income per day per household

What figure 4.9(next page) and Table 4.10(above) are reflecting is that majority of the rural population are earning less than a dollar a day with a mean of Ksh. 55 a day per household. There is family which can not raise a shilling a day but only (Ksh 0.625). This is a grave situation given the prevailed cost of living by then, which is determined by external forces. Therefore intervening programs are recommended to reverse the situation of the rural majority.

Distribution of daily income per household

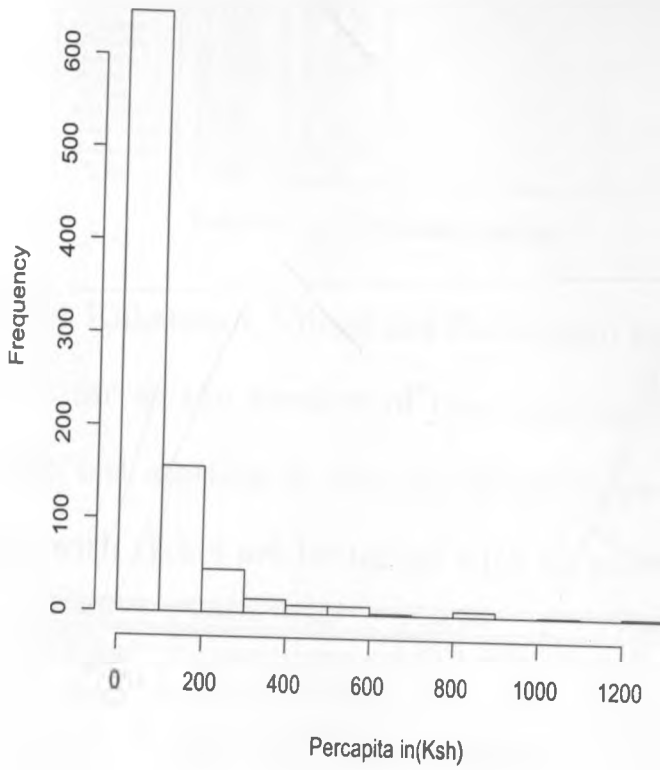


Figure 4.9: Daily household income

Variable	betteroff	poor	Test stat.	P value
District			Chisq. (14 df) = 92.58	0.001
Bungoma	6 (1.0)	31 (6)		
Kakamega	3 (0.8)	31 (6)		
Kiambu	51 (13.6)	53 (10.2)		
Kirinyaga	22 (5.9)	26 (5)		
Kisii	19 (5.1)	48 (9.2)		
Machakos	33 (8.8)	39 (7.5)		
Maragua	11 (2.9)	16 (3.1)		
Muranga	36 (9.6)	35 (6.7)		
Nakuru	81 (21.7)	75 (14.4)		
Nandi	28 (7.5)	34 (6.5)		
Narok	23 (6.1)	5 (1)		
Nyamira	7 (1.9)	26 (5)		
Nyandarua	31 (8.3)	19 (3.7)		
Rachuonyo	10 (4.3)	51 (9.8)		
Vihiga	7 (1.9)	31 (6)		

Table 4.11: Crosstabulations

Bungoma, Kakamega, Vihiga and Rachuonyo have the biggest disparity as far as the number of poor and better off is concerned. The test statistic is very significant since $0.001 \leq 0.05$ 42 of those with rights are better off with 67 poor. Those with

Variable	betteroff	poor	Test stat.	P value
Rights			Chisq. (2 df) = 2.77	0.25
Haverights	42 (11.2)	67 (12.9)		
No Rights	28 (7.5)	53 (10.2)		
Withapproval	304 (81.3)	400 (76.9)		

Table 4.12: Crosstabulations

no rights only 28 are better off with 53 poor. The test statistic is not statistically significant at 95% CI $0.25 \geq 0.05$

AgriKnowledge	betteroff	poor	Test stat.	P value
Official	67 (17.9)	69 (13.3)	Chisq. (2 df) = 5.49	0.064
Others	66 (17.6)	79 (15.2)		
Parents	241 (64.4)	372 (71.5)		

Table 4.13: Crosstabulations

Majority are getting the agricultural knowledge from their parents of which 241 are better off and 372 are poor. Only

136 are getting the knowledge from the formal source. The test statistic is not statistically significant at 95% CI $0.064 \geq 0.05$

4.4 Odds Ratio

4.4.1 Univariate Logistic Regression

Logistic regression predicting povindex : poor vs betteroff

Variable	Estimate	Std. Error	z value	OR	(95%CI)	Pr(> z)
(Intercept)	-	-	-	-	-	0.0015
sexmale	-0.1732	0.1652	-1.05	0.84	(0.61,1.16)	0.2944

Table 4.14: logistic analysis1

The findings show that the odds of females being poor is 1.19 times compared to males. The test statistic is not statistically significant at 95%CI since $0.2944 \geq 0.05$

province	Est	Std. Error	z-value	OR	(95%CI)	P(Wald's test)
Ref.=central	-	-	-	-	-	0.001
eastern	0.18039	0.26321	0.685	1.2	(0.71,2.01)	0.493
nyanza	1.10398	0.21247	5.196	3.02	(1.99,4.57)	0.001
riftvalley	-0.13327	0.17228	-0.774	0.88	(0.62,1.23)	0.439
western	1.77334	0.29426	6.027	5.89	(3.31,10.49)	0.001

Table 4.15: logistic analysis2

The results show that the odds of being poor in Eastern is 1.2 times in Central while in western is 5.89 times. The test statistic is significant at 95%CI

District	est	std.error	z-value	Or	(95%CI)	P(Wald's test)
ref.=kiambu	-	-	-	-	-	0.001
bungoma	1.60376	0.48724	3.292	4.97	(1.91,12.92)	0.001
kakamega	2.29691	0.63551	3.614	9.04	(2.86,34.55)	0.001
kirinyaga	0.12859	0.34985	0.368	1.14	(0.57,2.26)	0.713
kisii	0.88830	0.33458	2.655	2.43	(1.26,4.68)	0.008
machakos	0.12859	0.30728	0.418	1.14	(0.62,2.08)	0.676
maragua	0.33623	0.43805	0.768	1.4	(0.59,3.3)	0.443
Muranga	-0.06684	0.30794	-0.216	0.94	(0.51,1.71)	0.829
nakuru	-0.11543	0.25329	-0.456	0.89	(0.54,1.40)	0.649
nandi	0.15569	0.32187	0.484	1.17	(0.62,2.2)	0.629
narok	-1.56452	0.53099	-2.946	0.21	(0.07,0.59)	0.003
nyamira	1.27372	0.46882	2.717	3.57	(1.43,8.96)	0.007
nyandarua	-0.52801	0.35123	-1.503	0.59	(0.3,1.17)	0.133
rachuonyo	1.12077	0.34725	3.228	3.07	(1.55,6.06)	0.001
vihiga	1.44961	0.46210	3.137	4.26	(1.72,10.54)	0.002

Table 4.16: logistic analysis3

Kiambu is the reference. The results show that the odds of somebody in Bungoma being poor is 4.97 times compared to one in Kiambu. Odds of one in Vihiga being poor is 4.26 times compared to one in Kiambu. The test statistic is statistically significant in Vihiga, Rachuonyo, Nyamira, Narok, Kisii Bugoma and Kakamega.

House size	Est	Std. Error	z value	OR	(95%CI)	P(Wald's test)
ref.=1	-	-	-	-	-	0.001
2	-0.1102	0.0651	-0.175	0.89	(0.24,3.28)	0.861
3	1.3850	0.5478	2.528	3.99	(1.37,11.69)	0.011
4	1.6314	0.5345	3.052	5.11	(1.79,14.57)	0.002
5	1.8733	0.5251	3.568	6.51	(2.33,18.22)	0.001
6	1.7528	0.5255	3.336	5.77	(2.06,16.16)	0.001
7	2.2192	0.5329	4.164	9.2	(3.24,26.15)	0.001
8	1.9400	0.5349	3.627	6.96	(2.44,19.85)	0.001
9 Above	2.5597	0.5187	4.935	12.93	(4.68,35.74)	0.001

Table 4.17: logistic analysis4

	Estimate	Std. Error	z value	OR	(95%CI)	Pr(> z)
(Intercept)	-	-	-	-	-	0.2577
agegr30-44	-0.3489	0.6312	-0.553	0.71	(0.2,2.43)	0.5805
agegr45-54	-0.4418	0.6270	-0.70	0.64	(0.19,2.2)	0.4810
agegr55+	-0.3410	0.6193	-0.55	0.71	(0.21,2.39)	0.5819

Table 4.18: logistic analysis5

Source of Variance	Estimate	Std. Error	z value	OR	(95%CI)	P(Wald's test)
AgricalKnowledge	-	-	-	-	-	-
ref.=official	-	-	-	-	-	-
others	0.15038	0.23922	0.629	1.16	(0.73,1.86)	0.53
parents	0.40468	0.19041	2.125	1.5	(1.03,2.18)	0.034

Table 4.19: logistic analysis6

The odds of one being poor are increasing with an increase of household member. The test is statistically significant at 95%CI. For the age and agriculture knowledge source, the tests are not significant.

4.4.2 Multivariate logistic regression

```
fit4=glm(povindex~age+sex+hsizel+district +rights+
Agriculture.Knowledge,family=binomial(link=logit),data)
anova(fit4)
step(fit4,method="both")
summary(fit4)
logistic.display(fit4)
#running step(fit4,method="both")will give series of models.
#best fit is 5 because age has no influence when present.
# the lowest AIC 1096
fit5=glm(povindex~sex+hsizel+district +rights+AgricalKnowledge,
family=binomial(link=logit),data)
summary(fit5)
logistic.display(fit5)
```

	Estimate	Std. Error	z value	OR	(95%CI)	Pr(> z)
(Intercept)	0.3258	0.6065	0.54			0.5911
sexmale	-0.3605	0.1883	-1.91	0.84	(0.61,1.16)	0.0555
hsize1	0.2414	0.0356	6.79	1.27	(1.19,1.36)	0.0000
District: ref.=Bungoma						0.001
Kakamega	0.6976	0.7696	0.91	2.01	(0.44,9.08)	0.3647
Kiambu	-1.3612	0.5097	-2.67	0.26	(0.09,0.7)	0.0076
Kirinyaga	-0.9676	0.5490	-1.76	0.38	(0.13,1.11)	0.0780
Kisii	-0.6351	0.5391	-1.18	0.53	(0.18,1.52)	0.2388
Machakos	-1.1999	0.5197	-2.31	0.3	(0.11,0.83)	0.0210
Maragua	-0.8729	0.6120	-1.43	0.42	(0.13,1.39)	0.1538
Muranga	-1.2184	0.5273	-2.31	0.3	(0.11,0.83)	0.0208
Nakuru	-1.5335	0.4931	-3.11	0.22	(0.08,0.57)	0.0019
Nandi	-1.5980	0.5302	-3.01	0.2	(0.07,0.57)	0.0026
Narok	-3.4824	0.6836	-5.09	0.03	(0.01,0.12)	0.0000
Nyamira	-0.1230	0.6348	-0.19	0.88	(0.25,3.07)	0.8463
Nyandarua	-1.7656	0.5502	-3.21	0.17	(0.06,0.5)	0.0013
Rachuonyo	-0.4682	0.5485	-0.85	0.63	(0.21,1.83)	0.3933
Vihiga	0.0140	0.6249	0.02	1.01	(0.3,3.45)	0.9821
Rights: ref.=haverights						
No rights	0.1742	0.3397	0.51	1.19	(0.61,2.32)	0.6081
With approval	-0.3687	0.2340	-1.58	0.69	(0.44,1.09)	0.1151
AgricalKnowledge: ref.=official						
Others	-0.0143	0.2626	-0.05	0.99	(0.59,1.65)	0.9567
Parents	0.3587	0.2104	1.70	1.43	(0.95,2.16)	0.0883

Based on the multivariate age does not help in poverty prediction. Empowerment(rights), agriculture knowledge, gender help us to predict the level of poverty however they are not significant at 95%CI. The most significant predictors are the housesize and location(District). As per the sex the odds have not changed between univariate and multivariate ie chances of female being are 1.19 compared to their counterparts (males).The odds of Kakamega being poor is 2.01 compared to Bungoma and the odds of vihiga is 1.01. As far as the source of agricultural knowl-

edge is concerned the odds of those who get the knowledge from parent being poor is 1.43 compared to those officially trained. The odds of those without powers to make their own decisions of being poor is 1.19 compared to those with powers(rights).

Chapter 5

SUMMARY OF FINDINGS, CONCLUSION AND RECOMMENDATIONS

5.1 Summary of the findings

All the predictor variables we used displayed some relationship with the dependent variable, however; gender, age and source of agricultural knowledge were not significant at 95% CI under univariate logistic regression. Housesize, empowerment and location were very significant. After running our best model and

estimating the betas(coeffients) we came up with the equation of predicting the rural poor as shown below.

```
y<-as.numeric(povindex)
```

```
X<-cbind(sex,hsizel,district,rights,AgricalKnowledge)
```

```
A=t(X)%*%X
```

```
B=solve(A)
```

```
D=t(X)%*%y
```

```
Betas=B%*%D
```

```
Betas
```

	[,1]
sex	0.172177471
hsizel	0.079655440
district	0.002832943
rights	0.109674194
AgricalKnowledge	0.177126092

Hence our equation becomes

$$\Pi = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} = 0.32581 + 0.172sex + 0.08housesize + 0.003district + 0.11rights + 0.177agrknowledge + error\ term.$$

The results imply that source and gaining agricultural knowledge has the highest impact(0.177) on rural poverty reduction, followed by gender disparit(0.172) in that order upto the last which is location(0.003).

5.2 conclusion

The multivariate analysis shows that when a household increases by one member odds of poverty increases by 0.08, hence family planning is vital for rural poverty alleviation. It is important to realise districts with low potential for development intervention programs. Empowering people i.e giving people the power to make their own decision has a great impact also on poverty as shown (0.1097) and lastly I would wish to say agricultural knowledge is very essential for rural poverty reduction, Multivariate analysis gives it the highest coefficient (0.177) with

0.32581 as the intercept. The aim of this paper was to establish the demographic characteristic of the rural poor and to identify whether there is any statistically significant relationship between the predictors and the poverty level. We can conclude that the rural poor has a huge household size, no or limited rights, poor agricultural knowledge which not formal.

5.3 Recommendations

Suggestions for further research may include the use of insecurity, accessibility to financial facilities and asset ownership to predict the rural poverty level. There is need to come up with a better model which is all inclusive for poverty index determination.

Bibliography

- [1] Altman, D. G., Machin, D., Bryant, T. N. and Gardner, M. J. (2000) *Statistics with Confidence*, British Medical Journal, London.
- [2] Aquiar Mark and Erik Hurst (2004): *Consumption vs. Expenditure Working Paper*, University of Chicago.
- [3] Brown, C. C. (1982) On a goodness of fit test for the logistic model based on score statistics. *Communications in Statistics Theory and Methods*, 11, 1087-1105.
- [4] Charnes, A., Frome, E. L. and Yu, P. L. (1976) The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *Journal of the American Statistical Association*, 71, 169-171.

- [5] Chelliah R. J. and Sudarshan, R. (1999). Income - poverty and beyond. Human development in India. New Delhi: Esha Beteille Social Science Press
- [6] Collett, D. (1994) Modelling Binary Data. Chapman and Hall.
- [7] Cook, R. D. and Weisberg, S. (1999) Applied Regression including Computing and Graphics, Wiley, New York.
- [8] Diggle, P. J., Liang, K.-Y. and Zeger, S. L. (1994) Analysis of Longitudinal Data, Oxford University Press, Oxford.
- [9] Essenberg, J. M. (1952) Cigarette smoke and the incidence of primary neoplasm of the lung in albino mouse. Science, 116, 561 - 562.
- [10] Fahrmeir, L. and Kaufmann, H. (1985) Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. Annals of Statistics, 13, 342-368

- [11] Graybill, F. A. (1976) Theory and Application of the Linear Model, Duxbury, North Scituate, Massachusetts
- [12] Hoblyn, T. N. and Palmer, R. C. (1934) A complex experiment in the propagation of plum rootstocks from root cuttings: season 1931 - 1932. Journal of Pomology and Horticultural Science, 12, 36 - 56
- [13] Hogg, R. V. and Craig, A. T. (1995) Introduction to Mathematical Statistics, Fifth Edition, Prentice Hall, New Jersey.
- [14] Johnson David S. and Timothy M. Smeeding (1998b): Intergenerational Equity in the United States: The Changing Well-being of the Old and the Young, 1960-1995, Working Paper, Bureau of Labor Statistics.
- [15] kkasina United Nations Development Program (1997). Poverty measurement: Behind and beyond the poverty line (United Nations Development Program Technical Support Document Module 3). Washington, DC: Author.

- [16] McCullagh, P. (1980) Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B*, 42, 109-142.
- [17] Montgomery, D. C. and Peck, E. A. (1992) *Introduction to Linear Regression Analysis*, Second Edition, Wiley, New York
- [18] Nancy Khadioli (2008) Key determinants of poverty using asset index to get poverty index. University of Nairobi school of Mathematics.
- [19] National Centre for HIV Epidemiology and Clinical Research (1994) *Australian HIV Surveillance Report*, 10.
- [20] Paul Kamau (2006) *Gender disparity as a cause of poverty*. University of Nairobi, School of economics
- [21] Wood, C. L. (1978) Comparison of linear trends in binomial proportions. *Biometrics*, 34, 496-504.

Apendix A

R-Script

```
\singlespacing{  
\itshape{library(MASS)  
library(epicalc)  
library(xtable)  
\# read data from excel  
data=read.csv('martin.csv',header=T)  
\# analysis  
data\%$size2=as.factor(data\%$size1)  
attach(data)  
des(data)  
use(data)  
\# descriptive statistic  
summ(age)  
tab1(sex)  
summ(percapita)
```

```

tab1(district)
hist(perc capita,main='Distribution of daily income per
person',xlab='Percapita in(Ksh)')
tableStack(vars=c(9,10,25,5),by=povindex)
table=tableStack(vars=c(9,10,25,5),by=povindex)
xdesc=xtable(table)
tabpct(sex,povindex)
tableStack(rights,by=povindex)
tabpct(agegr,povindex)
\#test of relationship
tableStack(rights ,by=povindex)
tableStack(AgricalKnowledge ,by=povindex)
tableStack(vars=c(19,20),by=povindex)
xtep=tableStack(vars=c(19,20),by=povindex)
xtep=xtable(xtep)
print(xtep,floating=FALSE)
\# Bungoma ref
fit1=glm(povindex~district,family=binomial(link='logit'),data)
summary(fit1)

```



```

logistic.display(fit1)
\# as kiambu as the reference
data\${district}<-relevel(district,ref='kiambu')
fit2=glm(povindex~district,family=binomial(link='logit'),data)
summary(fit2)
fit21=xtable(fit2)
print(fit21,floating=F)
logistic.display(fit2)
fit211=logistic.display(fit2)
fit211=xtable(fit211)
print(fit211,floating=FALSE)
fit3=glm(povindex~hsize2,family=binomial(link='logit'),data)
summary(fit3)
logistic.display(fit3)
fit31=xtable(fit3)
print(fit31,floating=F)
fit4=glm(povindex\~age+sex+hsize1+district +rights+
AgriculKnowledge,family$=binomial(link=logit)$,data)
anova(fit4)

```

```

step(fit4,method=both)

summary(fit4)

logistic.display(fit4)

#fit5 is the best after running step(fit4,method=both)
#it has the lowest AIC and no variable age
#when in the model i.e doesn't make our it better.

fit5=glm(povindex~sex+hsize1+district +rights+AgriculKnowledge,
family$=binomial(link=logit)$,data)}}

y<-as.numeric(povindex)

X<-cbind(sex,hsize1,district,rights,AgriculKnowledge)

A=t(X)%*%X

B=solve(A)

D=t(X)%*%y

Betas=B%*%D

Betas

```