



UNIVERSITY OF NAIROBI

**Using Logistic regression in studying factors
associated with choice of savings mechanisms for
households in Kenya**

BY

Christine Anyango Okeyo

I56/38551/2020

A Thesis Submitted to the Department of Mathematics for Examination in Partial
Fulfillment of the Requirements for the Award of Degree of Master of Science in
Social Statistics of the University of Nairobi

November 2022

Using logistic regression in studying factors associated with choice of savings mechanisms for households in Kenya

Christine Okeyo

Department of Mathematics
Faculty of Science and Technology
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master Thesis

Submitted to the Department of Mathematics in partial fulfilment for a degree in Master of Science in Social Statistics

Submitted to: The Graduate School, University of Nairobi, Kenya

Abstract

Savings plays an important role in the lives of individuals in a country. Moreover, saving using formal means enables them to enjoy benefits such as safe keeping of money as well as easy access to other financial products such as loans. Despite the growth in access of financial services in Kenya, there is a segment of the population that doesn't make use of such savings products. This study explores the association between financial literacy, access and use of technology and socio-demographic variables on savings mechanisms of households in Kenya. The study used a secondary data source from the 2021 Finaccess Household Survey.

The main objectives of the study was to examine the association between financial literacy, access and use of technology and socio-demographic variables on use of formal savings as well to find a parsimonious model that best defines the relationship. Logistic regression was fitted to evaluate the association between the independent variables and households' use of formal savings mechanisms. Lasso model was then used to determine variables that were most predictive of using formal savings mechanisms as the most important saving device. A reduced model was then fitted based on the variables that were most predictive of the likelihood of households using formal savings as the most important saving mechanism.

The findings indicated that financial literacy, ownership of mobile phones, use of the internet, having disability, gender, highest education attained, religion, income source were significantly associated with use of formal savings as the most important savings mechanisms. The reduced model was significantly better at predicting the likelihood of saving formally as the main savings mechanism compared to the full model. The study recommended that relevant stakeholders should create financial products that are more inclusive towards certain segments of the population. They can also use the model created to identify individuals with the least likelihood of using formal savings methods as the most important one.

Declaration and Approval

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.



23/11/2022

Signature

Date

CHRISTINE ANYANGO OKEYO

Reg No. I56/38551/2020

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.



23/11/2022

Signature

Date

Dr. George Muhua
Department of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: muhugeorge@gmail.com

Dedication

I am dedicating this project to my mother Rosemary Okeyo and siblings Dorcas, Nereah and Fredrick Okeyo.

Contents

Abstract	ii
Declaration and Approval	iv
Dedication	vii
Acknowledgments	x
1 Chapter 1: Introduction	1
1.1 Introduction.....	1
1.2 Background.....	1
1.3 Problem Statement.....	2
1.4 Research Objectives.....	3
1.5 Study Hypothesis.....	3
1.6 Significance of Study.....	3
2 Chapter 2: Literature Review	5
2.1 Introduction.....	5
2.2 Effect of financial literacy, access and use of technology, socio-demographic variables on savings mechanisms.....	5
2.2.1 Financial literacy on the choice of savings mechanisms.....	5
2.2.2 Access and use of technology on the choice of savings mechanisms.....	6
2.2.3 Demographic variables on the choice of savings mechanisms.....	6
2.3 Summary of the Literature Review.....	7
3 Chapter 3: Research Methodology	8
3.1 Introduction.....	8
3.2 Study Design.....	8
3.3 Data source.....	8
3.4 Definition of variables used in the model.....	8
3.5 Logistic Regression.....	10
3.6 Statistical Significance of Predictors.....	11
3.6.1 Wald Statistic.....	11
3.6.2 Odds ratios with 95% CI.....	11
3.7 Evaluating Model Fit.....	11
3.8 Lasso Logistic Regression.....	12
3.9 The likelihood ratio test.....	12
4 Chapter 4: Data Analysis and Results	14
4.1 Introduction.....	14
4.2 Descriptive statistics.....	14
4.3 Inferential statistics.....	14
4.3.1 Evaluating Model fit.....	16

4.3.2 Lasso Logistic Model	16
4.3.3 Reduced Model	17
4.3.4 Comparison of the full vs reduced models	18
4.3.5 Interpreting model parameters	19
5 Chapter 5: Discussion, Conclusion and Recommendations	21
5.1 Introduction	21
5.2 Discussion	21
5.3 Conclusion	21
5.4 Recommendations	22
References	23
References	23

Acknowledgments

I am grateful to God for giving me life and good health.

Special thanks to my supervisor Dr, George Muhua for the guidance and support in ensuring a successful project.

I also extend appreciation to the Department of Mathematics - University of Nairobi for giving me an opportunity to expand knowledge and skills in the field of statistics as well as giving capable personnel to guide me through my academic life.

Lastly, I am very thankful to my mother Rosemary Okeyo for always giving me support and the strength to carry on during turbulent times.

Christine Okeyo

Nairobi, 2022.

1 Chapter 1: Introduction

1.1 Introduction

This chapter gives the background of study, the problem statement, research objectives, study hypothesis and significance of the study.

1.2 Background

Savings plays a vital role in the lives of individuals and households generally given that it helps to cushion them against the negative effect of shocks on their social and economic outcomes.

According to (Brune, Giné, Goldberg, & Yang, 2016), (Dupas & Robinson, 2013), (Karlan, Ratan, & Zinman, 2014), savings help families control their spending, increase their wealth, and making investments in human capital, such as health and education, which increases their ability to endure financial shocks.

In 2021, 76% of adults worldwide held an account with a bank or other regulated organization, such as a credit union, microfinance organization, or supplier of mobile money services. 31% of adults—or about two-thirds of people who saved any money reported to using formal means. (Demirguc-Kunt, Klapper, Singer, & Ansar, 2022)

In 2021, in Sub-Saharan Africa, 55% of adults possessed a formal account, including 33 percent of individuals who had a mobile money account. 26% of individuals — or about half of savers—saved formally. (Demirguc-Kunt et al., 2022)

According to (Demirguc-Kunt et al., 2022) in Kenya, in 2021, account ownership was at 79% and 45% of households saved formally or about two thirds of savers saved formally.

Looking at the 2030 Sustainable Development Goals, financial inclusion is considered to be a key driver behind eight of the seventeen goals. These goals touch on social aspects affecting individuals in our society today. Some of them include; eradicating poverty, ending hunger, attaining gender equality, promoting of sustainable agriculture and achieving food security. Additionally, the implicit role for greater financial inclusion through increased savings mobilization for investment and consumption that can stimulate growth is present in SDG 17 on enhancing the means of implementation. (UNCDF, 2019)

Formal savings in this study has been defined as savings in formal financial service providers such as banks, MFIs, SACCOs and Mobile Money while informal savings is defined as savings using informal providers such as giving to friends for safe keeping and saving in a secret hiding place.

Government officials and other key decision-makers have made tremendous efforts and advancements in the area of financial inclusion particularly with regard to household access to financial products. However, given that a portion of the population still employs informal savings methods, they miss out on the advantages of formal instruments. Additional research is needed to understand aspects that could increase adoption of formal financial products.

According to the 2021 Finaccess survey, 37% of people who save still hide their money in secret places or give it to their friends and family for safe keeping. Given the high rate of mobile money ownership in the nation and the fact that it has long been recognized as a key factor in the development of formal financial inclusion worldwide, these findings still show that more can be done in terms of encouraging households' use of formal savings instruments hence necessitating more research on factors associated with use of formal savings.

Most studies that have been done in Kenya on understanding households' savings behaviours have mostly explored the the casual relationship of demographic variables on choice of savings' instruments.

A study by (P. Morgan & Trinh, 2019) found that, while controlling for income and education, people with a high level of financial literacy scores were more likely to make use of formal savings than people with a low level of financial literacy.

1.3 Problem Statement

While financial inclusion has been generally increasing over time in Kenya, there is still a relatively higher percentage of households that don't make use of formal financial savings products hence saving money in informal savings platforms such as secret hiding places therefore being excluded in terms of savings instruments. According to the 2021 Finaccess household survey data, only 62% of the households reported to using formal savings instruments as the most important savings device. This remains a primary concern at the household level because there are households that do not fully reap the benefits of use of formal savings instruments such as security of their savings, interest growth on their principal deposit and access to credit using their savings as collateral.

The exclusion of these households, also remains a concern at the national level given that financial institutions use pooled resources to give out credit. Compared to the savings

rate, respondents' use of credit increased far more quickly between 2016 and 2021. This could mean that lenders need to increase their efforts to mobilize savings to meet the rate of lending demand or explore for new funding sources. (CBK, KNBS, & Kenya, 2021)

Multiple studies have been done on households use of formal savings techniques but there are several socio-demographic and use and access to technology variables that are yet to be explored. Moreover, the studies only identified the casual relationships of the various independent variables on use of formal savings techniques but have not explored finding a reduced statistical model that defines the relationship. This study seeks to fill that gap.

1.4 Research Objectives

The main purpose of this study is to evaluate factors associated with formal savings of households in Kenya

The specific objectives are:

1. To examine the association between use of formal savings mechanisms and socio-demographic factors, financial literacy and use of technology of households in Kenya
2. To devise a parsimonious model using financial literacy, access and use of technology and socio-demographic variables on households' use of formal savings mechanisms

1.5 Study Hypothesis

- $H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$; There is no association between socio-demographic factors, financial literacy and use of technology and formal savings mechanisms of households in Kenya
- $H_a : \beta_1 = \beta_2 = \dots = \beta_k \neq 0$; There is some association between socio-demographic factors, financial literacy and use of technology and formal savings mechanisms of households in Kenya
- H_0 : The fit of the full model and the reduced model is the same
- H_a : The full model fits the data significantly better than the reduced model

1.6 Significance of Study

The conclusions derived from the study may be used to bring change to policies that are associated with inclusion involving usage of formal savings products. The government and other stakeholders may use the findings to change some of the service delivery and outreach so as to ensure formal savings products that are designed to fit the needs of the people.

Additionally, formal financial institutions may use the outputs from this study to improve predictability of likelihood of use of formal savings products resulting in an improvement in their customer relations and improved targeting on the type of services provided which would in turn improve financial inclusivity and user experience. Moreover, future researchers as well as academicians may use the findings of this study to base their investigations and research in the subject of financial inclusion.

2 Chapter 2: Literature Review

2.1 Introduction

This chapter highlights a review of research on this topic of study. Review is done based on the different independent variables of interest - financial literacy, use and ownership of technology and socio-demographic variables.

2.2 Effect of financial literacy, access and use of technology, socio-demographic variables on savings mechanisms

2.2.1 Financial literacy on the choice of savings mechanisms

(Adetunji & David-West, 2019) conducted research on how financial inclusion in Nigeria is impacted by income and financial literacy. The results showed that savings habits with formal and informal savings institutions were significantly impacted by financial literacy. It was discovered that respondents with low financial literacy levels and low incomes were more likely to save at home or with friends. Additionally, it was discovered that formal savings were significantly influenced by both income and financial literacy levels, although income was not the only factor. The study made use of survey information from over 22,000 Nigerian respondents, which was statistically representative of the country's adult population. Savings frequency was utilized to measure financial inclusion. Three broad measures were used to calculate the frequency of saving: (i) the frequency of saving in banks and other formal deposit-taking institutions; (ii) the frequency of informal saving in cooperative societies and savings groups; and (iii) the frequency of saving in other contexts outside of formal and informal financial institutions (with family and friends and at home)

The independent variables used in the study were monthly income levels and financial literacy variable that was generated from questions measuring financial knowledge which were measured between a scale of 1 to 10. The effect of these variables of the outcome variable was then measured while controlling for the effects of age, gender and rural-urban classification variables.

Three ordered logistic regression models were used to identify the relationship between the independent variables and the three measures of formal savings. The three independent variables were classified using an ordered scale on the frequency of savings using

the different platforms. They were defined as; frequency of savings in formal institutions, frequency of savings in informal groups and frequency of savings at home or with friends.

(P. J. Morgan & Long, 2020) carried out a study on determinants and impacts of financial literacy in Cambodia and Viet Nam. The study found that both financial literacy and general education levels were found to be positively and significantly related to saving behaviour and financial inclusion.

Data was collected through a survey to respondents that would be representative of the adults in Cambodia and Viet Nam. The dependent variable used for savings was a binary variable; 1 for respondents that held some form of a savings product and 0 otherwise. The independent variables used was financial literacy comprised of variables on financial knowledge, attitudes and skill, income, education level, age, gender, source of income and rural/urban classification. A linear probability model and probit models were used to understand the effect of different independent variables on savings behaviour.

2.2.2 Access and use of technology on the choice of savings mechanisms

(Bayar, Gavriletea, & Păun, 2021) did a study to determine how internet use and mobile phone use affect financial inclusion. The main objective was to investigate this link using a sample of 11 post-communist nations in the European Union from 1996 to 2017. The study used panel cointegration and causality analysis to measure short- and long-term correlations.

The study found both positive and negative correlations between internet usage and access to financial markets as well as mobile phone and access to financial markets.

Mobile phone subscriptions had a favorable impact on both financial market access in Bulgaria, Croatia, and Hungary as well as access to financial institutions in nations like Latvia, Lithuania, Poland, and Slovenia. Additionally, a negative correlation between mobile phone subscriptions and access to financial institutions and financial markets in the Czech Republic and Poland was observed.

They concluded that they can boost access to financial institutions in Bulgaria, Croatia, Czech Republic, Hungary, and Poland as well as to financial markets in Latvia and Slovenia by increasing Internet usage.

2.2.3 Demographic variables on the choice of savings mechanisms

(Steiner, Giesbert, & Bendig, 2009) conducted research on the need for financial services in rural Ghana. A household survey that was conducted in Brakwa and Benin in the

Asikuma/Odoben/Brakwa area of the Central Region of Ghana in February 2008 provided the data for this study.

By using a multivariate probit model on data from a household survey in rural Ghana, they simultaneously assessed the factors influencing household demand for savings, loans, and insurances.

The findings were that financial service uptake positively affected by education level, asset endowment, and regular (formal) work status. This supported the widely held view that poorer households had a higher likelihood of being shut out of the formal financial system than better-off families.

2.3 Summary of the Literature Review

The studies mentioned above studied the effect of financial literacy on households use of formal savings mechanisms. One of them measured financial literacy using the knowledge dimension only but this study will add the attitudes dimension as well. It measured the effect of a number of explanatory variables on savings preference. We will seek to identify a model with as minimal number of explanatory variables as possible using binary logistic regression model below that can predict the likelihood of use of formal savings as the most important saving mechanism

3 Chapter 3: Research Methodology

3.1 Introduction

This chapter outlines the study design, the data source used, definition of variables used, description of logistic model, significance testing for predictors, model evaluation, description of Lasso model and comparison between full model and the reduced model.

3.2 Study Design

In this study, correlational research design will be employed. A correlational study involves collecting data on two or more variables for each individual in a sample without manipulation and working out the correlation coefficient (Bordens & Abbott, 2014). The purpose of correlational studies is to reveal relationships between naturally occurring variables through the use of correlational statistics

3.3 Data source

The 2021 FinAccess Household Survey, used a cross-sectional survey design at the household level, was the secondary source used for this study. The Survey involved Kenyans aged 16 years and above living within conventional households in Kenya but the analysis focused on respondents older than 18 years old.

The Kenya Household Master Sample Frame (K-HMSF), which was created from the 2019 Kenya Population and Housing Census, served as the source of the sample for the survey. The K-HMSF is made up of 10,000 clusters that were chosen using PPS from the 128,239 Enumeration Areas (EAs) that were produced for the 2019 Population and Housing Census. The sampling units were chosen using a multi-stage stratified cluster sampling process. EAs, households, and an eligible household member were the three groups of units sampled. The 92 strata in the K-HMSF were randomly chosen to make up the EAs, which served as the main sample units. Given that the master sample from which the sample was drawn had been drawn with probability proportional to size, the EAs were systematically chosen with equal probability. A systematic sampling technique was used to choose 18 households from the list of households enumerated during the 2019 KPHC for each EA. The survey was restricted to one eligible participant per selected household.

3.4 Definition of variables used in the model

Table 3.4.1. Variable definition

Variable	Variable definition	Measurement
Dependent variable		
Most important saving mechanism	Binary variable based on mechanism that is considered as the most important saving place	1=Formal 0=Informal
Independent variables		
Financial literacy		
Financial literacy score	A continuous variables showing the average score of 3 questions measuring financial knowledge and attitudes ; Respondent's understanding of transactions cost Respondent's understanding of interest rates Respondent's attitude towards betting as a good way of eraning income 1 - if correct 0 - if wrong	Continuous
Ownership and use of technology		
Mobile phone onwership	A categorical variable showing respondents' phone ownership status	1 = Respondent owns the phone 2 = Phone co-owned 3 = Dont own use someone else's 4 = Don't own or use
Internet usage	An ordinal variable showing frequency of respondents' internet use	1 = Daily 2 = Weekly 3= Monthly 4 = Less often 5= Never
Demographic Variables		
Rural-Urban Classification	Rural Urban Classification	1=Rural 2=Urban
Age	Respondent Age	1= > 55 2=18 - 25 3=26 - 35 4=36 - 45 5=46 - 55
Gender	Respondent Gender	1=Male 2=Female
Wealth Quintile	A ordinal variable calculating the wealth quintile of each respondent	1=Lowest 2=Second Lowest 3=Middle 4=Second Highest 5=Highest
Anyone with chronic disease	A categorical variable checking whether anyone in the household has a chronic disease	0= no 1=yes
Disability Status	A nominal categorical variable checking whether the respondent has any form of disability	0=no 1=Yes
Source of income	A categorical variable displaying the respondents' primary source of income	1= Farming 2= Employed 3= Casual Worker 4=Self employed 5=Support from family 6=Other
Religion	A categorical variable showing the religion of the respondent	1=Christian 2 = Islam 3= Other
Highest Educ	A nominal categorical variable showing the highest education attained by the respondent	1=None 2=Some Primary 3=Primary Completed 4=Some secondary 5=Secondary completed 7=Some technical 8=Technical completed 9= Some university 10 = University completed

3.5 Logistic Regression

Logistic regression is a statistical technique that is used when the response of interest has two outcomes. The model is usually fitted to identify whether there is a significant statistical relationship between a response variable and the different independent variables captured in a study.

The model operates under some assumptions which are:

- (i) Linearity: There should be a linear relationship between any continuous predictors and the logit of the outcome variable
- (ii) Multicollinearity: Predictor variables should not be too highly correlated
- (iii) Independence of errors: Cases of data should not be related

Logistic regression model enables one to identify the effect of explanatory variables $x_1, x_2, x_3, \dots, x_k$ and the outcome variable Y. The outcome variable Y in this case will have two categories. 1 if the category has the outcome of interests and 0 if not.

The logistic regression model is given by:

$$p(Y = 1|X) = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}} \quad (1)$$

Where β_0, \dots, β_k the model parameters, and x_1, \dots, x_k independent variables. Maximum likelihood method is used to derive model coefficients.

In our study we will have:

$$p(Y = 1|X) = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12} x_{12})}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12} x_{12})}} \quad (2)$$

Where;

x_1 =Financial literacy Score x_2 =Mobile phone ownership x_3 =Internet use x_4 =Rural - Urban Classification x_5 =Age x_6 =Gender x_7 =Wealth Quintile x_8 =Whether anyone in the household has a chronic disease x_9 =Whether the respondent has any form of disability x_{10} =Source of income x_{11} =Religion x_{12} =Highest education

The logistic model is converted into a linear model using a logit link function due to the model's non-linearity

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x + \beta_2x_2 + \dots + \beta_kx_k \quad (3)$$

, where $0 \leq p \leq 1$

Logistic regression calculates the probability of success over probability of failure. The results of the analysis are in the form of an odds ratio.

$$\frac{p(Y = 1|X)}{1 - p(Y = 1|X)} = e^{(\beta_0 + \beta_1x + \beta_2x_2 + \dots + \beta_kx_k)} \quad (4)$$

The odds ratio gives a comparison of the odds of outcome in one group relative to another.

OR>1 demonstrates that exposure is linked to a higher likelihood of getting the outcome.
OR<1 demonstrates that exposure is linked to a lower likelihood of getting the outcome.
OR=1 demonstrates that the odds of an outcome are unaffected by exposure.

3.6 Statistical Significance of Predictors

3.6.1 Wald Statistic

The Wald statistic was used to calculate the overall effect of any independent variables in our model. The Wald statistic test is given by:

$$Z = \frac{(\hat{\beta}_j)}{se(\hat{\beta}_j)} \quad (5)$$

The Wald statistics follows a Chi-square distribution with 1 degree of freedom.

3.6.2 Odds ratios with 95% CI

Odds ratio with 95% confidence interval (CI) can be used to evaluate how well each predictor contributes.

The hypothesis was described as: $H_0 : e^{\beta_0} = 1$ vs $H_1 : e^{\beta_1} = 1$
where $100(1 - \alpha)\%$ confidence interval was defined as: $e^{\beta_1 - (z_{\frac{\alpha}{2}} * se(\beta_1))}$; $e^{\beta_1 + (z_{\frac{\alpha}{2}} * se(\beta_1))}$

3.7 Evaluating Model Fit

Model fit was evaluated using residual deviance. The value was compared to the null deviance. The test statistic was a distributed chi-square with the degrees of freedom

calculated as the difference between the fitted model's degree of freedom and the null model's degree of freedom.

$$D(y) = -2l(\hat{\beta}; y) + 2l(\hat{\theta}; y) \quad (6)$$

where; $\hat{\beta}$ represents the fitted model of interest and $\hat{\theta}$ represents the saturated model

3.8 Lasso Logistic Regression

Lasso regression model can result in reduced models with few coefficients by adding a penalty equivalent to the absolute magnitude of the coefficients, some of which reduce to zero. More severe penalties result in coefficient values that are closer to zero, which is excellent for building models that are easier to understand. To support the findings from the full regression model, this model was employed for variable selection.

The penalty term in the log likelihood function can be defined as: $|\beta_1|$ where;

β_1 is a vector with the total number of components that match the predictors

Thus the value to be minimised in this case would be;

$$L + \lambda \sum \beta_1^2$$

(7)

The parameter λ is chosen so that the resulting model minimizes the out of sample error. Typically, grid search with cross-validation is used to determine the ideal value of λ .

3.9 The likelihood ratio test

Likelihood ratio test was used to compare the fit of full model and reduced model. The difference between the reduced model and the full model produced a goodness of fit index G , 2 statistic with k degrees of freedom.

$$G = \chi^2 = -2 \log \frac{\text{likelihood of reduced model}}{\text{likelihood of full model}} \quad (8)$$

This is a measure of how well all of the independent variables affect our response variable. If the p-value is less than 0.05, we will conclude that the full model is better than the reduced model.

4 Chapter 4: Data Analysis and Results

4.1 Introduction

This chapter highlights findings in terms of descriptive statistics, inferential statistics specifically on testing hypothesis using a logistic model, testing the fit of the model, getting a parsimonious model using Lasso and comparison between the reduced model and the full model.

4.2 Descriptive statistics

32% of respondents reported to using informal savings channels as their most important savings mechanisms. The distribution by location was 64% were from rural areas while 36% were from urban areas. Looking at whether there was anyone in the household with a chronic disease, 18% responded with an affirmative while 13% of the respondents reported to having some form of disability. Exploring the distribution by wealth quintile, 23% of the respondents belonged to the highest quintile, while 22% belonged to the second highest quintile and 16% belonged to the lowest quintile. Majority of the respondents were Christians at 89%. Just over half of the respondents interviewed were female (58%). Looking at variables capturing use of technology, 88% of the respondents reported to owning the mobile phone that they use while on internet use, 62% of the respondents reported to never using the internet while only 22% reported to using the internet on a daily basis. The average financial literacy score was 0.64 with the highest possible score being 1.

4.3 Inferential statistics

The full model was fitted to identify the association between the different independent variables and use of formal savings mechanisms. The hypothesis being tested was:

- $H_0 : \beta = \beta_2 = \dots = \beta_k = 0$; There is no association between socio-demographic factors, financial literacy and use of technology and formal savings mechanisms of households in Kenya
- $H_a : \beta = \beta_2 = \dots = \beta_k \neq 0$; There is some association between socio-demographic factors, financial literacy and use of technology and formal savings mechanisms of households in Kenya

Table 4.3.1. Binary Logistic Regression Model - Full Model

		Savings preference				
		Oddsratio	Estimate	Std.Error	Z.value	Pr(> z)
	(Intercept)	2.6887	0.9890	0.1157	8.551	<0.0001
1*Financial Literacy	Financial.lit.score	1.5725	0.4526	0.0736	6.146	< 0.0001
7*Access and Use of Technology	mobi.own.co.own	0.5661	-0.5690	0.1142	-4.979	< 0.0001
	mobi.own.dont.own.use.else	0.1647	-1.8198	0.0843	-21.262	<0.0001
	mobi.own.dont.own	0.0637	-2.7541	0.1371	-20.078	<0.0001
	internet.weekly	0.8470	-0.1660	0.0873	-1.850	0.0642
	internet.monthly	1.0809	0.0778	0.1986	0.392	0.6950
	internet.less.often	0.7105	-0.3418	0.1007	-3.394	0.0007
	internet.never	0.6959	-0.3526	0.0711	-5.101	<0.0001
27*Demographic variables	region.Urban	0.9441	-0.0576	0.0556	-1.036	0.3002
	age.18-25	0.8978	-0.1078	0.0774	-1.393	0.1635
	age.26-35	1.0664	0.0643	0.0685	0.938	0.3481
	age.36-45	0.9750	-0.0253	0.0697	-0.363	0.7166
	age.46-55	0.9652	-0.0354	0.0753	-0.470	0.6384
	chronic.Yes	0.8846	-0.1226	0.0535	-2.289	0.0220
	disability.Yes	0.8572	-0.1540	0.0617	-2.498	0.0125
	Gender.Female	0.5647	-0.5715	0.04324	-13.216	<0.0001
	educ.some.prim	1.0136	0.01355	0.0709	0.191	0.8483
	educ.prim	1.1259	0.1186	0.0760	1.562	0.1182
	educ.some.sec	1.1093	0.1038	0.0905	0.147	0.2514
	educ.sec	1.4589	0.3777	0.0870	4.342	<0.0001
	educ.some.tech	2.0516	0.7186	0.1855	3.874	< 0.0001
	educ.tech	1.7207	0.5427	0.1235	4.394	< 0.0001
	educ.some.uni	2.2493	0.8106	0.2165	3.745	0.0002
	educ.uni	1.7456	0.5571	0.1636	3.406	0.0006
	quintile.second.lowest	1.0380	0.0373	0.0675	0.553	0.5801
	quintile.middle	1.1866	0.1711	0.0692	2.474	0.0133
	quintile.second.highest	1.3652	0.3113	0.0760	4.100	<0.0001
	quintile.highest	1.4101	0.3436	0.0925	3.715	0.0002
	religion.islam	1.3263	0.2824	0.0753	3.749	0.0002
	religion.other	0.6372	-0.4507	0.1719	-2.621	0.0088
	inc.employed	1.5777	0.4560	0.0896	5.088	<0.0001
	inc.casual	1.1597	0.1481	0.05720	2.589	0.0096
	inc.self	1.0869	0.0833	0.0650	1.283	0.1996
	inc.family	1.0922	0.0882	0.0642	1.374	0.1693
	inc.other	1.0267	0.0264	0.1338	0.197	0.8435

4.3.1 Evaluating Model fit

We evaluated the model fit using the deviance statistic. This compared the performance of our model compared to a null model. The chi-square value was 2976.846 with 35 degrees of freedom with a p-value was 0.00 indicated that our model fit significantly better than the null model.

Table 4.3.2. Goodness of fit for the full model

#Df	Chisq	Pr(>Chisq)
35	2976.846	0.0000

4.3.2 Lasso Logistic Model

Lasso logistic model was then used to model all the variables on the likelihood of using formal savings as the most important savings instrument. Both the coefficients for urban-rural classification as well as age were penalised to zero. This further validated the results from the full model and further necessitated removing the two variables.

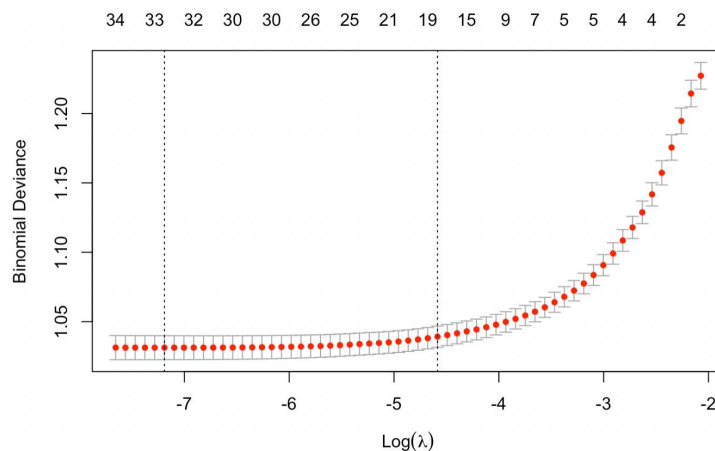


Figure 1. Lambda values selection

Penalised regression coefficients were obtained from lasso regression. It still supported the results from the full logistic model that both rural-urban classification variable and age were less predictive of likelihood to use formal savings as the most important saving technique.

```

##                                     s1
## (Intercept)                        1.06551829
## literacy1                          0.62743677
## mobi_ownI own a mobile phone that both me and others use -0.34699951
## mobi_ownI don't own a mobile phone but am able to use someone else's -1.66870436
## mobi_ownI don't own or use a mobile phone -2.37539167
## internet_useWeekly                  .
## internet_useMonthly                  .
## internet_useLess often               .
## internet_useNever                    -0.32114587
## RegionUrban                          .
## Age18 - 25                           .
## Age26 - 35                           0.06064448
## Age36 - 45                            .
## Age46 - 55                            .
## Chronic_diseaseYes                   -0.03474406
## DisabilityYes                         -0.10258134
## GenderFemale                          -0.47911920
## Highest_educSome primary              -0.06101230
## Highest_educPrimary completed         .
## Highest_educSome secondary            .
## Highest_educSecondary completed       0.04340730
## Highest_educSome technical training after secondary school 0.01102782
## Highest_educCompleted technical training after secondary school 0.11115765
## Highest_educSome university           .
## Highest_educUniversity completed      .
## quintileSecond Lowest                 -0.02311331
## quintileMiddle                         .
## quintileSecond Highest                0.10268920
## quintileHighest                       0.24326399
## ReligionIslam                         0.01539988
## ReligionOther Religion                 -0.20597798
## Income_sourceEmployed                  0.30944799
## Income_sourceCasual worker             .
## Income_sourceRunning own business/Self employed .
## Income_sourceMoney / support from family / friends / spouse .
## Income_sourceOther                     .

```

Figure 2. Penalised coefficients from Lasso Regression

4.3.3 Reduced Model

The reduced model was fitted without the urban-rural classification variable and the age variable that had been omitted.

Table 4.3.3. Binary Logistic Regression Model - Reduced Model

		Savings preference				
		Oddsratio	Estimate	Std.Error	Z.value	Pr(> z)
	(Intercept)	0.0450	-3.1019	0.3142	-9.8730	0.0000
1*Financial Literacy	Financial.lit.score	1.5736	0.45342	0.07337	6.180	0.0000
7*Access and Use of Technology	mobi.own.co.own	0.5628	-0.5748	0.1142	-5.034	0.0000
	mobi.own.dont.own.use.else	0.1621	-1.8198	0.0843	-21.569	0.0000
	mobi.own.dont.own	0.0629	-2.7661	0.1369	-20.192	0.0000
	internet.weekly	0.8465	-0.1666	0.0896	-1.859	0.0630
	internet.monthly	1.0791	0.0761	0.1983	0.384	0.7012
	internet.less.often	0.7137	-0.3373	0.1005	-3.355	0.0007
	internet.never	0.7026	-0.3529	0.0700	-5.041	0.0000
22*Demographic variables	chronic.Yes	0.8846	-0.1226	0.0525	-2.333	0.0197
	disability.Yes	0.8583	-0.152787	0.059895	-2.551	0.0107
	Gender.Female	0.5663	-0.568659	0.043049	-13.210	0.0000
	educ.some.prim	1.0098	0.0097	0.0702	0.139	0.8893
	educ.prim	1.1172	0.110871	0.074404	1.490	0.1361
	educ.some.sec	1.0845	0.0811	0.0885	0.917	0.3591
	educ.sec	1.4288	0.3569	0.0852	4.190	0.0000
	educ.some.tech	1.9649	0.6754	0.1836	3.679	0.0002
	educ.tech	1.7252	-0.0658	0.1608	-0.4100	0.6822
	educ.some.uni	2.1335	0.5453	0.1226	4.446	0.0000
	educ.uni	1.7600	0.5653	0.1631	3.465	0.0005
	quintile.second.lowest	1.0352	0.0346	0.0668	0.518	0.6045
	quintile.middle	1.1782	0.1641	0.0675	2.431	0.0150
	quintile.second.highest	1.3375	0.2908	0.0712	4.081	0.0000
	quintile.highest	1.3577	0.3057	0.0809	3.776	0.0002
	religion.islam	1.3045	0.2657	0.0741	3.586	0.0003
	religion.other	0.6348	-0.4544	0.1718	-2.645	0.0082
	inc.employed	1.5681	0.449	0.0889	5.055	0.0000
	inc.casual	1.1469	0.1370	0.0554	2.471	0.0135
	inc.self	1.0832	0.0798	0.0638	1.251	0.2108
	inc.family	1.0713	0.0688	0.0628	1.096	0.2729
	inc.other	1.0265	0.0261	0.1316	0.198	0.8426

4.3.4 Comparison of the full vs reduced models

Likelihood ratio test was ran to compare the full model and the reduced model. The output shows that the chi-square statistic is 17.34 and the corresponding p-value is 0.1253. Since the p-value is greater than 0.05, we fail to reject the null hypothesis and conclude that the reduced model fits the data equally well as the reduced model. See table 5

Table 4.3.4. Likelihood ratio test

#Df	LogLik	Df	Chisq	Pr(>Chisq)
37	-8089.8			
32	-8094.2	-5	8.6186	0.1253

From the regression model outputs we devise the following equation,

$$\begin{aligned}
 \ln\left(\frac{p}{1-p}\right) = & 0.665571 - 0.087115\textit{literacy} - 0.131382\textit{mobi.own.co.own} - 0.409806\textit{mobi.own.dont.own.use.else} \\
 & - 0.526585\textit{mobi.own.dont.own} - 0.016902\textit{internet.weekly} + 0.020665\textit{internet.monthly} - \\
 & 0.045204\textit{internet.less.often} - 0.052783\textit{internet.never} - 0.023077\textit{chronic.disease.yes} \\
 & - 0.030361\textit{disability} - 0.096609\textit{female} + 0.003632\textit{educ.some.pri} + 0.026596\textit{educ.pri.comp} + \\
 & 0.021750\textit{educ.some.sec} + \textit{educ.sec.comp} + 0.106923\textit{educ.some.tech} + 0.089547\textit{educ.comp.tech} \\
 & + 0.111352\textit{educ.some.uni} + 0.081843\textit{educ.comp.uni} + 0.004976\textit{quintile.sec.lowest} \\
 & + 0.030994\textit{quintile.middle} + 0.053597\textit{quintile.sec.high} + \\
 & 0.054309\textit{quintile.highest} + 0.046736\textit{religion.islam} - 0.076880\textit{religion.other} + \\
 & 0.062039\textit{inc.employed} + 0.026104\textit{inc.casual} + 0.016403\textit{inc.self} + \\
 & 0.012948\textit{inc.family} + 0.007608\textit{inc.other}
 \end{aligned} \tag{9}$$

4.3.5 Interpreting model parameters

When holding all variables constant, financial literacy score and use of technology have a significant statistical effect on the choice of savings mechanisms at 5% significance level.

While holding other variables constant, financial literacy is positively associated with use of formal savings as the most important savings instrument. An increase in one unit of financial literacy increases the likelihood of using formal savings as the main savings instrument by 9%.

Respondents who co-own a mobile phone i.e both them and others use are 23% less likely to use formal savings as the most important savings mechanisms while respondents that don't own a mobile phone but are able to use someone else's were 33.5% less likely and respondents who don't own or use a mobile phone were 41% less likely to use formal savings as the most important savings mechanism.

Holding all other factors constant, respondents that never use the internet were 6% less likely to use formal savings mechanisms as the main savings platforms compared to respondents that use internet daily while respondents that use the internet less often were 5% less likely and finally respondents that use the internet weekly were 2% less likely to use formal savings as the main savings instrument

Respondents in households with someone who has a chronic disease were 3% less likely to save formally compared to households without such an individual.

Holding all variables constant, disability was correlated with the likelihood of using formal savings.

Respondents with a form of disability were 24% less likely to use formal savings as compared to respondents without disability.

Wealth quintile was positively associated with the likelihood of saving formally. Respondents in the second lowest quintile were 0.4% more likely to save formally, ones in the middle were 3.1% more likely while second highest were 5.50% more likely while the highest were 5.58% more likely compared to the respondents in the lowest quintile.

While holding other variables constant, women were 46% less likely to use formal savings as the most important savings platforms compared to men.

While holding other variables constant, education level was somewhat positively correlated with likelihood of using formal savings. Respondents with higher education levels were more likely to use formal savings platforms as the main savings instrument compared to respondents with lower education levels. A respondent with some university education compared to a respondent with none was 2.13 times more likely to use formal savings as the main savings instrument compared to a respondent with no education.

5 Chapter 5: Discussion, Conclusion and Recommendations

5.1 Introduction

This section will conclude of the findings for the study, conclusions of this research as well as gaps to be filled by other research.

5.2 Discussion

The study found a statistically significant association between financial literacy and the likelihood to use formal savings as the most important savings mechanisms. This is in line with the findings from (Adetunji & David-West, 2019). These results showed that financial literacy plays a vital role in determining use of formal savings.

The statistically significant association between mobile ownership and likelihood to use formal savings as the most important savings mechanisms as well as internet use and likelihood to use formal savings as the most important savings mechanisms shows the how important access and use of technology is in addressing the gaps in financial inclusion from the demand side in a developing economy context that with a widespread use of mobile money access. This was in line with the findings from (Bayar et al., 2021)

More socio-demographic factors such as such as having disability, gender, religion of an individual were also found to have an effect on likelihood to use formal savings as the most important savings mechanism.

The parsimonious model created helped to identify the least number of predictors that were important important in predicting likelihood of using formal savings as the most important mechanism. As a result, narrowing down on important variables that can be used to understand formal savings in developing economies.

5.3 Conclusion

The study established that Financial literacy, access and use of technology, some socio-economic variables such as gender, disability status, wealth quintile, religion and source of income were significantly associated with the use of formal savings mechanisms.

While controlling for the effects of other factors, region and age did have a significant statistical association with the choice of savings mechanisms

5.4 Recommendations

From the findings, the government and relevant stakeholders should create empowerment programs that are aimed at enhancing financial literacy for the adult population in Kenya.

The study also recommends that financial products created should that cater for the needs of individuals who do not have primary ownership of their mobile phones as well as individuals with disabilities.

The government should provide an enabling environment for easy and affordable access to internet connectivity within the country that has shown to enhance use of formal savings mechanisms.

The created model can be used by relevant industry players in identifying individuals with the least likelihood of using formal savings as the most important mechanisms and design products and interventions that cater for their needs.

Further research should be done on exploring the frequency of use of different formal techniques. This enables to further understand usage of formal techniques in Kenya.

References

- Adetunji, O. M., & David-West, O. (2019, May). The Relative Impact of Income and Financial Literacy on Financial Inclusion in Nigeria. *Journal of International Development*, 31(4), 312–335.
- Bayar, Y., Gavriletea, M. D., & Păun, D. (2021, May). Impact of mobile phones and internet use on financial inclusion: empirical evidence from the EU post-communist countries. *Technological and Economic Development of Economy*, 27, 722–741.
- Bordens, K., & Abbott, B. B. (2014). *Ebook: Research Design and Methods: A Process Approach*. McGraw Hill.
- Brune, L., Giné, X., Goldberg, J., & Yang, D. (2016, January). Facilitating Savings for Agriculture: Field Experimental Evidence from Malawi. *Economic Development and Cultural Change*, 64(2), 187–220.
- CBK, KNBS, & Kenya, F. (2021). *2021 FinAccess Household Survey on financial inclusion*.
- Demirguc-Kunt, A., Klapper, L., Singer, D., & Ansar, S. (2022, June). *The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19* (Tech. Rep.). Washington, DC: World Bank. doi: 10.1596/978-1-4648-1897-4
- Dupas, P., & Robinson, J. (2013, June). Why Don't the Poor Save More? Evidence from Health Savings Experiments. *American Economic Review*, 103(4), 1138–1171.
- Karlan, D., Ratan, A. L., & Zinman, J. (2014). Savings by and for the Poor: A Research Review and Agenda. *Review of Income and Wealth*, 60(1), 36–78.
- Morgan, P., & Trinh, L. (2019, January). Determinants and Impacts of Financial Literacy in Cambodia and Viet Nam. *Journal of Risk and Financial Management*, 12(1), 19.
- Morgan, P. J., & Long, T. Q. (2020). Financial literacy, financial inclusion, and savings behavior in Laos. *Journal of Asian Economics*, 68(C).
- Steiner, S., Giesbert, L., & Bendig, M. (2009, January). *Savings, Credit and Insurance: Household Demand for Formal Financial Services in Rural Ghana* [SSRN Scholarly Paper]. Rochester, NY.
- UNCDF. (2019). *Financial Inclusion and the SDGs - UN Capital Development Fund (UNCDF)*.