University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master Project in Social Statistics

Determinants of Health Insurance Uptake Among Women in Kenya: An Application of Discriminant Analysis

Research Report in Mathematics, Number 26, 2017

GITHINJI REUBEN THUITA

December 2017

Submitted to the School of Mathematics in partial fulfillment for a degree in Master of Science in Social Statistics

Master Project in Social Statistics **University of Nairobi**
December 2017

Determinants of Health Insurance Uptake Among Women in Kenya: An Application of Discriminant Analysis

Research Report in Mathematics, Number 26, 2017

GITHINJI REUBEN THUITA

Master Project

Submitted to the School of Mathematics in partial fulfillment for a degree in Master of Science in Social Statistics

Prepared for The Director
Board Postgraduate Studies
University of Nairobi

Monitored by Director, School of Mathematics

Abstract

The current high utilization of out of pocket payments by a majority of Kenyans to settle their medical bills has continued to ensure that the poor and the vulnerable in the society cannot access essential health care services. Studies have shown that having a health insurance cover can greatly reduce the over-reliance on out of pocket financing. Despite studies showing that Kenya's population under health insurance coverage has grown, the population of women with health insurance schemes has continued to fall below the national average. The goal of this study was to examine determinants of health insurance uptake among women in Kenya using the discriminant analysis approach. The study used data from KDHS collected in the year 2014.After fitting a discriminant analysis model using the step-wise procedure, all the eight predictor variables, namely age,marital status, education level, employment,wealth quintile,place of residence,household size and access to media were found to be significant in discriminating as to whether a woman was insured or uninsured. The classification accuracy of the discriminant model was 86.9 per cent, and the model was found to be statistically significant and hence using the eight predictor variables, one can be able to classify a woman as to whether she is insured or uninsured.

Master Thesis in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

I the undersigned declare that this research project is my original work and has not been presented for a degree to any other institution.

________________ ________________

Signature Date

GITHINJI REUBEN THUITA
Reg No. I56/81237/2015

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

________________ ________________

Signature Date

PROF. J.A.M OTIENO
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: joseph_otieno@uonbi.ac.ke

Dedication

I dedicate this research project to my spouse Phyllis and my sons Eddy and Ivan for their moral support throughout the writing down of this project.

Contents

1 INTRODUCTION

1.1 Background of the study

The concept of health insurance has been widely used in many nations around the world to cater for related medical expenses either through privately purchased insurance, social welfare programs or social insurance. Before the advent of health insurance in the mid-20th century, patients were required to meet their medical expenses out of their pockets. However, due to the financial burden that this had on a majority of the population, it led to the introduction of public health insurance programs and later the private health insurance programs. Today, many nations around the world continue to face significant challenges in their implementation of health services delivery frameworks. According to Carrin and Waelkens (2005), one of these challenges is inadequate health care funding. In developed nations, for instance, 36 per cent of health-care financing is out of pocket. However, this is far better than in developing countries which stands at an average of 83 per cent (WHO, 2010). Dependence on out-of-pocket settlement of medical bills prevents people from seeking health care when required and those who do suffer the problem of financial impoverishment (Kutzin, 2013). As such, it is from this acknowledgment that developed and developing countries have come up with social health protection mechanisms to aid in addressing challenges that arise in their quest for the provision of health care services to their citizens. One of these mechanisms is the social health insurance (SHI) which consists of pooling resources into a common fund where these funds are then utilized to cater for health care expenses for the members (Aspalter et al. 2017). This is in addition to the privately purchased health insurance schemes.

Health insurance history in Kenya dates back to the 1960s when the National Hospital Insurance Fund (NHIF) was established in 1966 through an Act of Parliament. Today, health insurance schemes have evolved and are accessible through any of the three schemes, namely; community-based health insurance (CBHI), private insurance firms and the public health insurance scheme under the National Hospital Insurance Fund (NHIF). Community-based health insurance in Kenya was established in 1999 though it has very limited coverage,benefiting only a partly 170000 (Mathauer et al., 2008). Members in these schemes make small but regular contributions to these schemes which are used for settling their medical bills. CBHI in Kenya are controlled by KCBHFA. Other insurance schemes available in Kenya are the private health schemes as well as the social health insurance scheme under the NHIF. NHIF was established after independence in 1966 and it is a statutory scheme for all employees in the formal sector while people working in the informal sectors of the economy contribute to this scheme on a voluntary basis. The employed are required to pay statutory premiums towards this scheme based on their gross monthly salary, the amount ranging from a low of Kshs.150 to Kshs.1700.

1.2 Statement of the problem

The national health insurance coverage has increased from a low of 10 per cent in 2010 (WHO, 2010) to an average of 20 per cent in 2014 (World Bank, 2014). The problem is that while the national health insurance coverage has continued to grow, women still experience a lower coverage, at 7 per cent in 2010 (KNBS, 2010) compared to the national average of 10 per cent (WHO, 2010). This low uptake of health insurance for women is evident despite the overall country's population consisting of 50.2 per cent women according to the Kenya population census in 2009 (KNBS,2010). While literature points outs the various demographic and social-economic factors such as age, marital status, household size, education, access to media, place of residence, occupation and wealth quintile associated with health insurance uptake among individuals, health insurance uptake among women remains an understudied topic with a majority of studies focusing on the overall population. Similarly, a number of the available literature has used logistic regression in trying to answer this question.With this in mind, there is need to identify determinants of health insurance uptake among women in Kenya.

1.3 Objective of the study

1.3.1 Main objective

The overall objective of this study is to examine determinants of health insurance uptake among women in Kenya.

1.3.2 Specific Objectives

- To identify demographic and social economic factors which are significantly associated with health insurance uptake among women in Kenya using discriminant analysis.

- To determine the accuracy ratio of the resulting discriminant model.

1.4 Justification of the study

There is a need for great analytical tools that will aid policymakers so that the right policies can be implemented based on the findings generated through the application of these tools. As such, a great statistical model is helpful in identifying these factors with the aim ensuring that the right policies are put in place by the relevant authorities.

2 LITERATURE REVIEW

2.1 Introduction

This chapter reviews health care financing in Kenya and factors which are believed to affect health insurance uptake among women. The review is outlined in three sections, namely; health care financing in Kenya, demographic factors and social-economic factors.

2.2 Health care financing in Kenya

Financing health care in Kenya continues to be a great challenge thus hindering access to quality health care services to a majority of the population (WHO, 2010). In a bid to ensure that the Kenya achieves millennium development goals related to healthcare and reduce over-reliance on out of pocket expenditures to pay for health services, the country has adopted mixed health financing schemes. These schemes are namely; community-based health insurance (CBHI), private insurance firms and the public health insurance scheme under the National Hospital Insurance Fund (NHIF). The NHIF provides health insurance coverage to both formal and informal sector segments of the population. Formal sector persons pay income pegged premiums deducted monthly by the employer and remitted directly to the NHIF. These premiums are mandatory for all formal sector employees, and they range between Kshs.150 to Kshs.1700.The informal sector workers are required to pay a flat rate of Kshs.500 monthly on a voluntary basis. Over the years, the government of Kenya has carried out reforms within the NHIF to cater for both inpatient and outpatient services. On the other hand, private health insurance schemes are available in Kenya through privately owned institutions. Members join these insurance schemes on a voluntary basis with some employers paying for the schemes for their employees. CBHI schemes in Kenya have a voluntary membership and are operated under the KCBHFA. These schemes are based on the concept of social solidarity and mutual aid and are designed for the informal sector segment of the population. KCBHFA has five different benefit packages segmented in accordance with the amount of premiums paid by members. Despite the existence of the aforementioned health financing schemes in Kenya and the significant attempts by the government to reform the NHIF, a huge proportion of the population still remains uncovered with a majority of the population, 83 percent still relying on out of pocket payments to settle their medical bills (World Bank, 2014). According to WHO (2010), only 10 of the population in Kenya has some form of health insurance scheme. In particular, according to KNBS (2010) only 11 percent of men and 7 percent of women in Kenya between the ages of 15 to 49 years have an insurance cover. According to a study by the World Bank, only 20 percent of Kenyans have access to any form of health insurance

(World Bank, 2014). This situation has caused a large proportion of the population to be excluded from access to quality health care. Various studies have come up with a number of demographic and social economic factors have been found to be associated with health insurance uptake in various studies. These factors are discussed as below.

2.3 Demographic factors

2.3.1 Age

Various Studies have found out that there is a significant association between age and health insurance uptake. According to Mulenga et al. (2016) in a study to identify demographic and socio-economic determinants of women's health insurance coverage in Zambia using Logistic regression, the study revealed that there is a positive relationship between health insurance coverage and age. This is observation is consistent with a study conducted by Kiplagat et al. (2013) on determinants of health insurance choice in Kenya using multinomial logit model. Kiplagat et al. (2013) found that health insurance coverage increases with an increase in age. On the other hand, Mathur et al. (2014) in a study seeking to understand perception and factors influencing voluntary subscription to private health insurance using Logistic regression, it was observed that the mean age of the insured was much lower than the mean age of the uninsured. As such, the study observed that as age increased, the likelihood of purchasing health insurance decreased.

2.3.2 Marital Status

Researchers have shown that being married is associated with health insurance ownership. According to Kimani et al. (2014) in a study to identify determinants of health insurance ownership among women in Kenya using logistic regression found out that, there is a relationship between being married and having a health insurance cover as compared to formerly married and never being in a union. This is supported by studies conducted by Mulenga et al. (2016), Yue and Zou (2014) and Amu Dickson (2016). In a study conducted by Kirigia et al. (2005) on determinants of health insurance ownership among South African women using logistic regression model, the study showed that marital status had a statistically significant effect on the enrolment to health insurance. In this study, it was observed that persons who were married had a greater chance of having a health insurance cover as compared to their unmarried counterparts. This is further supported by a study carried out by Finn and Harmon (2006) which observed never married persons have a lower propensity to have an insurance cover as compared to the married or persons living with partners.

2.3.3 Place of residence

There are contradicting study findings regarding the association between an individual's place of residence and health insurance uptake. According to Kirigia et al. (2005) persons living in urban areas were seven times more likely to join a health insurance scheme as opposed to their counterparts living in rural areas. These findings are supported by a study conducted by Kimani et al. (2014).In addition, Mulenga et al. (2016) found out that 0.8 percent of women living in rural areas were covered by health insurance as opposed to 5.4 percent of their counterparts in urban settlements. On the other hand, Yue and Zou (2014) in a study where he used a bivariate probit analysis to examine the role of wealth and health in insurance choice in , it was observed that rural populations preferred health insurance as compared to their urban counterparts citing the implementation of rural cooperative medical scheme.

2.4 Socio-economic factors

2.4.1 Wealth quintile

Various studies have shown there is a significant positive relationship between the level of health insurance enrollment and income level. A study on determinants of health insurance uptake among women in South Africa by Kirigia et al (2005), it was observed that the proportion of women with health insurance increased as one's level of income increased. They found out that health insurance coverage stood at 6.3 percent for persons earning between Rand 1 to 950 while the coverage was at over 90.7 percent for those earning above Rand 7600 a month. Makoka et al. (2007) in a study on demand for private health insurance using multinomial logit model, it was found out that enrollment to various health insurance schemes was positively associated with an individual's monthly income. Additionally, this finding is supported by vast literature on determinants of health insurance coverage such as Liu and Chen (2002) as well as Finn and Harmon (2006).

2.4.2 Employment

Studies have shown that being in employment increases the odds of people enrolling in health insurance schemes. According to Butler (1999), persons who were gainfully employed were found to have higher chances of purchasing a health insurance cover. This study is consistent with the study conducted by Mulenga et al. (2016) which observed that women in employment had 70 percent higher odds of joining a health insurance schemes compared to the unemployed.

2.4.3 Household size

Various studies have contradicting observations regarding the relationship between enrolment to health insurance and household size. According to Onemolease Oriakhi (2012) in a study to identify determinants of rural household's willingness to participate in community based health Insurance scheme in Edo State in Nigeria using logistic regression, it was found that there is a positive association between health insurance purchase household sizes. This is to say that smaller families are less likely to purchase health insurance cover compared to larger families. Kirigia et al. (2005) on the other hand found out that there exists a negative relationship between household size and health insurance uptake.

2.4.4 Education level

Various studies argue that there is a positive association between education level and memberships to a health insurance schemes. In a study by Kiplagat et al. (2013), having attained tertiary education, secondary education and primary education was found to increase the likelihoods of purchasing a private health insurance cover by 22.5,4.2 and 3 times respectively as compared to having no education. This study is supported by a study carried out by Mulenga et al. (2016), which found out those women who had attained secondary education and higher had 510 percent higher chances of securing health insurance covers as compared to their counter parts. These observations are supported by studies conducted by Kimani et al. (2014) as well as Amu Dickson (2016).

2.4.5 Access to mass media

Studies have shown that access to information related to health insurance through the media increases the chance of joining a health insurance scheme. Kimani et al. (2014) found out that persons who read newspapers, listen to the radio, read newspapers and watch television almost on a daily basis have higher probability of 35, 58, and 19 percent respectively of joining a health insurance scheme. Likewise, in a study conducted by Mulenga et al. (2016) , the study revealed that a higher proportion of women who had access to mass media had health insurance coverage, this proportion being 3.7 percent as compared to their counterparts which stood at 0.5 percent.

2.5 Conceptual Framework

In order to examine the factors associated with the uptake of health insurance among Kenyan women, a framework representing the relationship between the dependent and predictor variables was developed as shown in figure 1. The development of this framework is based on available literature on the subject.

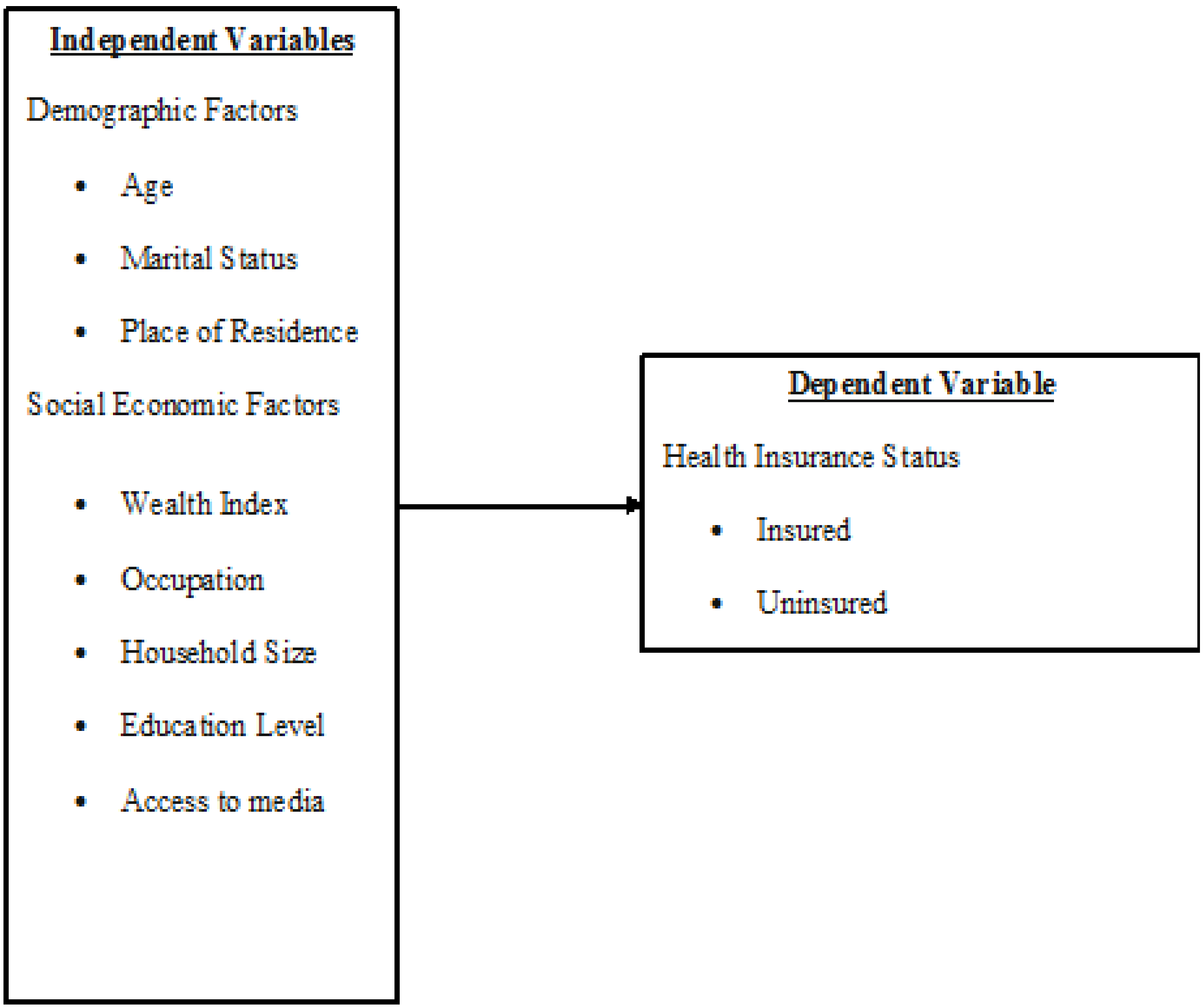

Figure 1: Conceptual Framework

3 METHODOLOGY

3.1 Introduction

The goal of this study was to identify significant determinants of health insurance uptake among women in Kenya. In order to ensure robust estimates for prediction purposes, it is vital to select the right analytical tool. However, apart from choosing the right analytical technique it is vital to apply a methodology that not only classifies a new case to its respective group but can also predict an outcome. Previous studies on the determinants of health insurance uptake have focused on the use of the multinomial logistic model and logistic regression analysis. While logistic regression analysis is used to analyze data with categorical dependent variables, discriminant analysis can also be applied for the same purpose. However, the choice between the two techniques depends on the underlying assumptions in each method. While logistic regression is termed as a distribution free test, linear discriminant analysis assumes normality on the continuous independent variables. However, according to Johnson and Wichern (2007), Fisher's linear discriminant analysis, a form of discriminant analysis makes no an assumption regarding normality of the population under study. Horeover, despite the dissimilarities between these two models concerning their assumptions, results of both techniques are analogous irrespective of the violation in the normality assumption.This happens especially when the sample size is large say 20 observations or more Antonogeorgos et al., (2009) and Pohar et al., (2004). With the information as described above, this study uses the Fishers Linear Discriminant Analysis to identify significant determinants of health insurance uptake among women in Kenya.

3.2 Data source

This study uses data collected in the year 2014 by the Kenya Demographic and Health Survey (KDHS). This data contains information on a sample of 14733 women. Information relating to their socio-economic and demographic characteristics was collected using a questionnaire. For purposes of validation, the sample has been split twice, 70 percent of the dataset was used as the analysis set while 30 percent was used as the holdout sample in order to test for the predictive accuracy of the resulting discriminant model.

3.3 Variable description

According to Hair et al., (2010), independent variables chosen in Fisher's linear discriminant analysis model can be two or more but they must be distinct and unique as well as be mutually exclusive and exhaustive. According to Uddin et al., (2013), the explanatory variables can be both metric and non-metric. For this study, in order to achieve the objectives of predicting health insurance uptake among women, the outcome variable which is categorical has been converted into a dummy variable with 0 and 1 representing uninsured and insured respectively. The explanatory variables, in this case, were place of residence, age, marital status, wealth quintile, occupation, household size, access to media and education level.
Since most of the predictor variables are categorical in nature, they have been converted to dummy variables.

3.4 Analytical method

For the purpose of this study, we used Fisher's Linear Discriminant Analysis methodology. This methodology is a statistical classification technique developed by Fishers in 1936 to classify observations to their respective groups based on a set of explanatory variables. The aim of discriminant analysis is to find a linear combination of predictor variables that best discriminants between groups. Hence, this methodology is used to predict group membership of a case from a given set of independent variables. This technique requires that we have one categorical grouping variable which is the outcome variable and two or more explanatory variables that are either categorical and/or continuous in nature.

3.4.1 Linear Discriminant function

Fisher's linear discriminant analysis encompasses having a linear combination which predicts group membership. This model takes the form

$Z_{ij} = \alpha + C_1X_k + C_2X_k + \ldots + C_nX_k$ .... (1)
Where:
Z_{ij} is the mean discriminant score for discriminant function i for object j
α is the model's discriminant constant
C_i is the discriminant coefficient for explanatory variable i also referred to as discriminant weight
X_k is the explanatory variable k.
Discriminant analysis technique like linear regression uses the ordinary least squares technique to estimate the values of the model parameters α and C_i that minimizes the

within-group sum of squares. After making these estimates, we then solve for Z_{ij} for each case.Solving for Z_{ij} for a particular case helps us in obtaining a score that is useful in predicting group membership for that case.

3.5 Assumptions of discriminant analysis

3.5.1 Multi-collinearity

Discriminant analysis assumes that there must be no correlation between explanatory variables. Presence of such correlation between various predictors would lead to bias in the analysis. Presence of strong multicollinearity between the various predictors means that we cannot be able to totally discriminate among groups. This is to imply that, when there is multicollinearity, such variables are said to be redundant and hence they do not add any value on how to separate groups. The presence of Multicollinearity can be checked by the use of stepwise discriminanat process when fitting the model or the correlation matrix. By using stepwise discriminant analysis, the problem of multicollinearity is addressed as only the significant variables are included in the model using this process.

3.5.2 Equal variance–covariance matrix

Discriminant analysis assumes that the variance – covariance matrix of explanatory variables are equal within each group of the dependent variable. This assumption is tested using Box's M test. A p value associated with the Box's M statistic implies that the assumption is violated and that the variance- covariance structure are statistically different. Under this test, the null hypothesis states that, the variance-covariance matrices of the two groups are equal. However, incase this assumption is violated it can often be ignored since discriminant analysis is robust for large sample sizes.

3.5.3 Outliers

Discriminant analysis is extremely susceptible to the presence of outliers. This is because these outliers have great influence the mean, statistical significance and standard deviation. The existence of these extreme observations can be tested by the use of mahalanobis distance. Those cases that are found to have large values of mahalanobis distance from their group means are regarded as outliers. It is important to eliminate outliers to get reasonable results.

3.5.4 Deriving the Coefficients of Fishers Linear Discriminant Model

To estimate the model's parameters in this study, we used Fisher's linear discriminant analysis. This technique seeks is to find discriminant coefficients such that the resulting model maximizes the distance between the two classes or groups. Fisher's procedure for classifying cases to their groups was to transform the multivariate observations X_1, X_2, X_3,..., X_q to a uni-variate variable Y which is a linear combination of the X's such that;

Y = $C_1X_1 + C_2X_2 + \ldots + C_qX_q$ .... (2)

This transformation is in such a way that the Y values resulting from the two groups are separated as much as possible.Assuming that we have values y_{11},y_{12},y_{13},...,y_{1n1} and y_{21},y_{22},y_{23},...,y_{2n2} for groups one and two respectively which are obtained from a fixed linear combination of the X's, then the separation between these two sets of values is defined as the absolute difference between their means expressed in standard deviation units.This is defined as;

$$Separation = \frac{|\bar{y}_1 - \bar{y}_2|}{s_y}$$

Where

s_y^2= $(\sum_{j=1}^{n_1}(\bar{y}_{1j} - \bar{y}_1)^2) + \sum_{j=1}^{n_2}(\bar{y}_{2j} - \bar{y}_2)^2)/n_1 + n_2 - 2$ the pooled variance estimate.

The goal is to choose a linear combination of the X's so that the separation obtained between the sample means $\bar{y}_1$ and $\bar{y}_2$ is maximized.

Suppose that we re-write model (2) as follows

y=$C_0 + \vec{C}^T\vec{X}$ 3

Then, the aim is to find a vector $\vec{C}^T$ such that if we project the predictors data along it, then the two sample means are separated as far as possible with the variances remaining as close as possible. Thus, we can think of the best $\vec{C}^T$ as the direction along which the two groups are well separated. Supposing that we have two data sets which are univariate, we can define their variances and means as follows;

$\exp(\hat{Y}/X \in g_i)$ = $Ci + C^T\mu_i$

$Var(\hat{Y}/X \in g_i)$= $C^T\mu_iC$

In this case, we find C that maximizes the ratio

$J(C)$=$\frac{(\mu_1^T C-\mu_2^T C)^2}{C^T S_1 C+C^T S_2 C}$.... (4)

The ratio J(C) above is the difference between the projected means of two groups normalized by a measure of within class variability

Implying that

J(C)=$\frac{[(\mu_1^T-\mu_2^T)C]^2}{C^T(S_1+S_2)C}$ ...(5)

Define

$S_p=S_1+S_2$ and L = $\mu_1^T-\mu_2^T$

Replacing L and $S_p$in equation (5) above, we have

$J(C)$=$\frac{(L^T C)^2}{C^T S_p C}$... (6)

Defining S_p as follows $S_p=M^T M$, where M is the square root of S_p equation (6) becomes

$J(C)$ =$\frac{(L^T C)^2}{C^T M^T M C}$... (7)

Projecting C through M and create a vector U such that
$U=MC$, hence $C=M^{-1}U$...(8)
Replacing C in equation (7) with equation (8) we obtain

$J(C)$=$([(M^{-1})^T R]^T \frac{U}{|U|}])^2$ .... (9)
Now, we need to find a vector U that will maximize equation (9) by projecting it whilst ensuring that the two vectors U and $([(M^{-1})^T L]^T$ are in the same direction. The vector U that maximizes equation (9) is given as;

U $=a([(M^{-1})^T L]^T$... (10)
Where a, is a constant.
Replacing L and U in equation (10), we obtain
$MC = a(M^{-1})^T(\mu_1 - \mu_2)$
Implying that, $C = a(M^T M)^{-1}(\mu_1 - \mu_2)$
Hence; $C = S^{-1}(\mu_1 - \mu_2)$... (11)
Where S is referred to as the pooled within group matrix.
Therefore, equation (11) gives us the Fisher's linear discriminant coefficients as desired.

3.6 Determining the goodness of fit and significance of the discriminant model

Once we have estimated the model coefficients and fitted the discriminant model, it is important to determine the goodness of fit and the statistical significance of the fitted model. This can be done using different statistical criteria namely; Eigen value, canonical correlation and the Wilk's lambda.

3.6.1 Eigen values

Eigen value λ is the ratio of explained variation to the unexplained variation in the fitted model. The Eigen value assesses the discriminatory power of the model. An Eigen value greater than 1 indicates a good model. It is computed as;
$$\lambda = \frac{BSS}{WSS} = \frac{\sum(\bar{Z}_j - \bar{Z})^2}{\sum(Z_{ij} - \bar{Z}_j)^2}$$

Where BSS refers to the between-group sum of squares, WSS is the within-group sum of squares. Noting that in discriminant analysis the total sum of squares (TSS) is partitioned into BSS and WSS such that;

$$TSS = \sum(Z_i - \bar{Z})^2 = BSS + WSS = \sum(\bar{Z}_j - \bar{Z})^2 + \sum(Z_{ij} - \bar{Z}_j)^2$$
Where: i is an individual case
j refers to the j^{th} group
Z_i is the discriminant score of the i^{th} individual
$\bar{Z}$ is the overall mean of the discriminant scores
Z_{ij} is the discriminant score of the i^{th} individual in group j
$\bar{Z}_j$ is the mean discriminant score in group j.

3.6.2 Canonical correlation

Canonical correlation (η) indicates the correlation between the predictor variables with the discriminant scores obtained from the model. The square of the canonical correlation indicates the proportion of variation in the dependent variable (discriminant scores) explained by the variations in the predictors. Canonical correlation is thus computed as;
$\eta = \sqrt{BSS/TSS} = \frac{\sum(\bar{Z}_j - \bar{Z})^2)}{\sum(Z_i - \bar{Z})^2} = \sqrt{\frac{\lambda}{1+\lambda}}$

3.6.3 Wilk's lambda

The Wilk's lambda Λ tests the significance of the discriminant model. The smaller the lambda is for an independent variable, the more that predictor variable contributes to the discriminant model. It is computed as; $\Lambda = \frac{WSS}{TSS} = \frac{\sum(\bar{Z}_j - \bar{Z})^2}{\sum(Z_i - \bar{Z})^2)} = \frac{1}{1+\lambda} = 1 - \eta^2$.
Wilk's lambda can be converted to a chi square statistics with k-1 degrees of freedom, where k is the number of parameters estimated in the model. This statistic is computed as $[(n-1) - 0.5(m+k+1)]ln\Lambda$.Where n is the combined number of cases from the two groups, m is the number of discriminant functions extracted which is equal to the number of groups minus one and k is the number of predictor variables.In the chi-square test, the null hypothesis is that the mean discriminant scores of the groups are equal.

3.7 Determining how well the model predicts

Once we have tested the statistical significance of the model, we need to determine how well the model predicts. To do this, we need to construct a classification matrix. The classification matrix is defined as the cross-tabulation of observed group membership and the predicted group membership. The numbers of the correctly classified cases are found on the leading diagonal while the numbers that are off-diagonal represent incorrectly classified cases. The overall percentage of the correctly classified cases is referred to as the hit ratio. For classification purposes where we have two groups, we use the cutting score to determine which group a case belongs.
For equal group sizes, the cutting score is the mean score, that is;
$Z_{cuttingscore} = \frac{\bar{z}_1 + \bar{z}_2}{2}$

Where $\bar{Z}_1$ is the mean discriminant score (centroid) for group 1
$\bar{Z}_2$ is the mean discriminant score for group 2
For unequal group sizes, the cutting score is computed from weighted means, such that
$Z_{cuttingscore} = \frac{n_1\bar{Z}_1 + n_1\bar{Z}_2}{n_1 + n_2}$

Where
n_1 is the number of observations in group 1 n_2 is the number of observations in group 2
To test whether the model's hit ratio is significantly better than chance we employ two tests namely; t-test for equal group sizes and the Press's-Q statistic for the unequal group sizes. The null hypothesis to be tested in this case is that the model hit ratio is no better than chance.

3.7.1 T-test for equal group sizes

When the two groups have equal sizes, i.e. $n_1 = n_2$ we employ a t-test. The statistic is computed as;

t= $(p-0.5)/\sqrt{0.5/N}$
with N-2 degrees of freedom Where: N is the total number of cases, P is the proportion of the cases correctly classified in the model.

3.7.2 Press's Q statistics for a model with unequal group sizes

When the group sizes are unequal, we use the Press's Q statistic. The Press's Q statistic follows a chi-square distribution with 1 degree of freedom. The statistic is computed as;
Q=$\dfrac{(N-n*g)^2}{N*(g-1)}$
Where: N is the total number of cases, n is the number of cases which are correctly classified, g is the number of groups. If the value of Q exceeds the critical value, we conclude that the model's hit ratio is significantly better than chance.

4 DATA ANALYSIS AND RESULTS

4.1 Descriptive statistics

4.1.1 Group statistics

Uninsured

insurance status	mean	std dev	valid N	valid N
			unweighted	weighted
age	28.45	9.446	8735	8735
hshold size	5.55	2.535	8735	8735
place of residence	0.66	0.475	8735	8735
education	0.84	0.362	8735	8735
access to media	0.80	0.402	8735	8735
wealth quintile	0.52	0.500	8735	8735
marital status	0.71	0.54	8735	8735

Insured

insurance status	mean	std dev	valid N	valid N
			unweighted	weighted
age	31.73	8.540	1578	1575
hshold size	4.67	2.388	1578	1578
place of residence	0.47	0.499	1578	1578
education	0.98	0.129	1578	1578
access to media	0.97	0.165	1578	1578
wealth quintile	0.89	0.500	1578	1578
marital status	0.80	0.403	1578	1578

Total

insurance status	mean	std dev.	valid N	valid N
			unweighted	weighted
age	28.95	9.387	10313	10313
hshld size	5.41	2.533	10313	10313
place of residence	0.63	0.483	10313	10313
education	0.87	0.341	10313	10313
access to media	0.82	0.381	10313	10313
wealth quintile	0.57	0.494	10313	10313
marital status	0.72	0.447	10313	10313
occupation	0.17	0.378	10313	10313

Table 4.1 above provides the descriptive statistics for the two groups, the insured and uninsured as well as the overall analysis sample.

4.2 Box's M test of equality of covariance matrices

Table 4.2: Test of equality of covariance matrices using Box's M

Insurance uptake	Rank	Log	Box's M	Approx. F	sig.
uninsured	8	-2.544	499.934	13.862	0.000
Insured	8	-2.971			
pooled within groups	8	-2.561			

In discriminant analysis, one of the assumptions is that of homogeneity of covariance matrices. To test this assumption, we employ the Box's M test whereby we seek a non – significant M for the assumption to hold. A non – significant M is equivalent to having the log-determinant values being very close to each other. The hypothesis to be tested in this case is;

H0: the covariance matrices of the two groups are homogeneous.

From our data, the assumption of equal covariance matrices is violated since M is found to be significant, $p < 0.05$. However, discriminant analysis is a robust technique for large samples ($n > 20$) even when this assumption is not met, and hence the violation, in this case can be ignored.

4.3 Fishers Linear Discriminant Function

The discriminant function is fitted by computing unstandardized canonical discriminant function coefficients. These coefficients are used to construct the actual prediction model that can be used in classifying new cases. Besides, we also have standardized canonical discriminant function coefficients; these standardized coefficients are used in understanding the discriminatory power of each predictor variable. The higher the absolute value of the standardized discriminant coefficient is for a certain predictor variable, the higher is the discriminatory power of that variable.
For this study, Fisher's linear discriminant function was obtained. From the data, using stepwise discriminant function analysis procedure, all the eight predictor variables were found to statistically significant. The stepwise procedure ensures that only significant variables are entered in the model. Using these eight predictor variables with their corresponding unstandardized discriminant function coefficients, the required Fisher's linear discriminant function becomes,

$$Z = -3.073 + 0.119age + 0.108place \quad of \quad residence + 0.761education \breve{} 0.025 \quad householdsize + 0.191access \quad tomedia + 0.280wealth \quad quintile + 0.455maritalstatus + 0.257employment ...$$
(12)

Using the above discriminant function (12) one can now be able to classify a Kenyan woman as to whether she has a health insurance cover or not based on the above set of predictors. Table 4.3 below shows a summary of interpretative measures of the discriminant analysis output.

Table 4.3: Summary of Fisher's discriminant analysis output

predictor variable	unstandardized	standardized	Wilks lambda	sig.
age	0.119	0.221	0.822	0.000
place of residence	0.108	0.052	0.807	0.000
education	0.761	0.576	0.876	0.000
hshld size	-0.025	-0.064	0.808	0.000
access to media	0.191	0.175	0.808	0.000
wealth quintile	0.280	0.378	0.845	0.000
marital status	0.455	0.203	0.811	0.000
employment	0.257	0.179	0.815	0.000
group centroid(insured)		1.149		
group centroid (uninsured)		-0.208		
Wilk's lambda		0.807		
canonical correlation		0.439		

4.4 Statistical significance of the discriminant model

To check if the discriminant model is statistically significant we use the model's Wilk's lambda derived from the eigen value. The Wilk's lambda follows a chi-square distribution with k − 1 degrees of freedom where k is the number of parameters estimated. The model's Wilk's lambda is shown in table 4.4 below.

Table 4.4: Statistical significance of the model

Function	Eigen Value	Canonical Correlation	Wilk's lambda	chi-sq	sig
1	0.239	0.439	0.807	2205.876	0.000

Thus from the table above, the hypothesis to be tested in this case to check if the model is statistically significant is;
H0 : In the population $\bar{Z}_1 = \bar{Z}_2 = \bar{Z}$ From the results in table 4 above, since the chi-square results are significant at 0.05 level, we reject H0 and conclude that the differences in the mean discriminant scores of the two groups are greater than what could be attributed to sampling error and hence the model is a good fit for the data and can be used for classifying new cases into the two distinct groups.

4.5 Discriminant criterion for classifying new cases

To identify a discriminant criterion for classifying new cases into either insured or uninsured, we use the group centroids. These group centroids are compared to the discriminant scores for a particular case to determine which group the case belongs to. From the results in table 4.3, the centroids for the uninsured are -0.208 while that of the insured is 1.149. While classifying new cases, if the discriminant score of a case is negative, then that case is likely to be grouped as uninsured, and if positive it is classified as insured.

4.6 Assessing the predictive accuracy of the model

It is important to note that in discriminant analysis, significances tests performed on the discriminant function are not sufficient to be able to make sound conclusions. This is because, with large samples, we can still achieve a small centroid distance that is significant. To deal with this, we need to assess the internal validity of the model. The internal validity of a discriminant model is assessed through the construction of a classification matrix using the holdout sample. The holdout sample produces the best accuracy rate (hit ratio) since it is not used in deriving the discriminant function. To check if the model predicts better than chance, three benchmarks are used, i.e., the maximum chance criterion, proportional chance, and the Press Q statistic.

Table 4.5: Hit ratio for cases used in the analysis sample

		predicted grou membership	predicted group membership
actual group	number of cases		
		uninsured(per cent)	insured(per cent)
uninsured	8735	8456(96.8)	279(3.2)
insured	1578	1060(67.2)	518(32.8)

87.0 per cent of the original grouped cases correctly classified

Table 4.6: Hit ratio of the holdout sample

		predicted group membership	predicted group membership
actual group	number of cases		
		Uninsured(per cent)	Insured(per cent)
Uninsured	3759	3650(97.1)	109(2.9)
Insured	661	471(71.3)	190(28.7)

86.9 per cent of the original grouped cases correctly classified

Computing the three benchmarks as mentioned above we obtain;

$Maximum \quad chance = \dfrac{3759}{4420} = 0.85$

$Proportional chance = P^2 + (1-P)^2 = (0.85)^2 + (0.5)^2 = 0.75$ Press Q statistic, $Q = \dfrac{(N - n * g)^2}{N * (g-1)}$

$= \dfrac{(4420 - 3840 * 2)^2}{4420} = 2404.43$

The three benchmarks computed above are presented in summary in table 4.7 below.

Table 4.7: Goodness of fit results as compared to chance

Benchmark	Value	Hit ratio (holdout sample)
Maximum chance	0.85	86.9
Proportional chance	0.75	86.9
Press Q calculated value	2404.43	
Press Q table value	6.635	

From the results in table 4.5 above, the predictive accuracy of the model from the analysis sample is 87.0 per cent, and the holdout sample hit ratio is 86.9 per cent as shown in table 4.6. The holdout sample hit ratio of 86.9 per cent as shown in table 4.7 above exceeds both the maximum chance and proportional chance values at 85 per cent and 75 per cent respectively. Besides, the Press Q statistic calculated value of 2404.43 exceeds the table value, 6.635 implying that it is significant. Based on the three benchmarks as highlighted above, we reject the null hypothesis that the model's hit ratio is no better than chance and conclude that classification, as derived from the model is significantly better than chance.

5 CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

This study examined the effects of demographic, social economic characteristics and access to media on the uptake of health insurance among women in Kenya using Fisher's linear discriminant analysis. All the eight variables highlighted in the study were found to be significant and hence included in the fitted discriminant model using stepwise discriminant analysis procedure. The discriminant model fitted was statistically significant based on statistical indicators including maximum chance criterion, proportional chance, and Press Q statistic. As such, the discriminant model so constructed can be used to classify a woman in Kenya as to either uninsured or insured based on the eight predictor variables.

5.2 Recommendations

To help understand determinants of health insurance uptake among Kenyan women, other factors not limited to individual demographics and social-economic characteristics need to be studied. These factors may include health insurance policy formulations by policy makers. Based on the findings from this study, policies should be aimed at enhancing those factors that are positively associated with health insurance uptake among women in Kenya as well as working on those factors that are negatively associated with health insurance coverage.

Bibliography

[AD] Amu, H., Dickson, K. (2016). *Health insurance subscription among women in reproductive age in Ghana: do socio-demographics matter?.* Health Economics Review, 6(1).

[an09] Antonogeorgos, G., Panagiotakos, D., Priftis, K., Tzonou, A. (2009). *Logistic Regression and Linear Discriminant Analyses in Evaluating Factors Associated with Asthma Prevalence among 10- to 12-Years-Old Children: Divergence and Similarity of the Two Statistical Methods. International Journal Of Pediatrics*, 2009, 1-6.

[AS17] Aspalter, C., Pribadi, K., Gauld, R. (2017). *Health care systems in developing countries in Asia.* Milton Park, Abingdon, Oxon: Routledge, Taylor, and Francis.

[But99] Butler, J. (1999). *Estimating Elasticities of Demand for Private Health Insurance in Australia. National Centre For Epidemiology And Population Health, 42.*

[car et al.05] Carrin, G., Waelkens, M., Criel, B. (2005). *Community-based health insurance in developing countries: a study of its contribution to the performance of health financing systems.* Tropical Medicine And International Health, 10(8), 799-811.

[FiHa] Finn, C., Harmon, C. (2006). *A Dynamic Model of Demand for Private Health Insurance in Ireland.* Institute For The Study Of Labor, (2472), 38.

[GOB16] Gobir, A., Adeyemi, A., Abubakar, A., Audu, O., Joshua, I. (2016). *Determinants of Willingness to Join Community- Based Health Insurance Scheme in a Rural Community of North-Western Nigeria.* African Journal Of Health Economics, (2), 9.

[Ha07] Hair, J., Black, W., Babin, B., Anderson, R. (2010).*Multivariate data analysis (7th ed.). Upper Saddle River, N.J.: Pearson/Prentice-Hall.*

[RW07] Johnson, R. A., Wichern, D. W. (2007). *Applied multivariate statistical analysis. Upper Saddle River, NJ: Pearson/Prentice Hall.*

[KNBS] Kenya National Bureau of Statistics (KNBS). (2010). *Economic Survey 2010.Nairobi*: Government Printer.

[KIM] Kimani, J., Ettarh, R., Warren, C., Bellows, B. (2014). *Determinants of health insurance ownership among women in Kenya: evidence from the 2008–09 Kenya demographic and health survey.* International Journal For Equity In Health, 13(1), 27.

[KIP13] Kiplagat, I., Muriithi, M., Kioko, U. (2013). *Determinants of Health Insurance Choice in Kenya*. European Scientific Journal, 9(13), 16.

[Sa05] , J., Sambo, L., Nganda, B., Mwabu, G., Chatora, R., Mwase, T. (2005). *Determinants of health insurance ownership among South African women*. BMC Health Services Research, 5(1).

[KU13] Kutzin, J. (2013). *Health financing for universal coverage and health system performance: concepts and implications for policy*. Bulletin Of The World Health Organization, 91(8), 602-611.

[LIU] Liu, T., Chen, C. (2002). *An analysis of private health insurance purchasing decisions with national health insurance in Taiwan*. Social Science Medicine, 55(5), 755-774.

[MAK] Makoka.,D,Kalua,.B Kambewa,.P (2007). *Demand for Private Health Insurance Where Public Health Services are Free: The Case of Malawi*. Journal Of Applied Sciences, 7(21), 3268- 3273.

[MAT] Mathauer, I., Schmidt, J., Wenyaa, M. (2008). *Extending social health insurance to the informal sector in Kenya. An assessment of factors affecting demand*. The International Journal Of Health Planning And Management, 23(1), 51-68.

[MAP] Mathur, T., Paul, U., Prasad, H., Das, S. (2014). *Understanding perception and factors influencing private voluntary health insurance policy subscription in the Lucknow region*. International Journal Of Health Policy And Management, 4(2), 75-83.

[MUL] Mulenga, J., Bwalya, B., Gebremeskel, Y. (2016). *Demographic and socio-economic determinants of women's health insurance coverage in Zambi*a. Epidemiology Biostatistics And Public Health, 14(1), 9.

[ORIA12] Oriakhi H. O, Onemolease E. A., (2012). *Determinants of Rural Household's Willingness to Participate in Community Based Health Insurance Scheme in Edo State, Nigeria*. Journal Of Ethnobiology And Ethnomedicine, 6(2), 17.

[POH04] Pohar, M., Blas, M., Turk, S. (2004). *Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study*. Metodološki Zvezki, 1(1), 143-161.

[UD13] Uddin , N., Meah, S. M., Hossain, R. (2013). *Discriminant analysis as an aid to human resources selection and human resources turnover minimization decisions. International Journal of Business and Management*, 8(17).

[WHO] WHO. (2010). *Health systems financing: the path to universal coverage*. World Health Organization.

[WB14] World BankWorld Bank. (2014). *Improving health outcomes and services for Kenyans: sustainable institutions and financing for universal health coverage - Kenya Health Policy Forum*. Documents.worldbank.org.

[YUE] Yue, Y., Zou, J. (2014). *The Role of Wealth and Health in Insurance Choice: Bivariate Probit Analysis in China.* Mathematical Problems In Engineering, 2014.