University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master Project in social statistics

Application of Logistic Regressions to Test Equality of Resource Allocation for Malaria Control in Kenya

Research Project in Mathematics, Number 20, 2017

Dominic I. Sambuli

August 2017

Submitted to the School of Mathematics in partial fulfillment for a degree in Master of Science in social statistics

Application of Logistic Regressions to Test Equality of Resource Allocation for Malaria Control in Kenya

Research Project in Mathematics, Number 20, 2017

Dominic I. Sambuli

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

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 of Postgraduate Studies
University of Nairobi

Monitored by Director, School of Mathematics

Abstract

Kenya is one of the developing countries that are hit by malaria cases and its prevalence being experienced more in some regions than others. Researchers and medical organizations have classified the country into five zones (highlands, lake basin, coastal, semi-arid and low risk zones) based on the fever cases recorded. Various strategies and objectives have been formulated with the aim of eliminating the disease from the nation. The laid objectives were based on demarcations of the country into the five zones.

Unfortunately, it has been difficult to eradicate the disease. This may be due to the fact that, most resources reserved for prevention and treatment of the fever are injected in areas mapped as Malaria endemic zones only; no serious action is taken in regions marked as non-endemic [WHO15].

This paper will try to test whether it is valid to have disparity in allocating resources for treatment and control of malaria across the country. It will use regression models and odd ratios.
It will also focus on certain groups of people in the society who are prone to the fever, and whether the preventive and treatment measures are done in a way that is proportional to the demand in each region. The conclusion will be advices to financiers, donors and governments when allocating resources geared towards achievement of millennium development goal of eliminating the disease in Kenya.

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

Signature	Date

Dominic Ingosi Sambuli
REG. NO. I56/81877/2015

In my capacity a supervisor of the candidate's project, I certify that his work has my approval for submission.

Signature	Date

Ms Anne Wang'ombe
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: awangombe@gmail.com

Dedication

This project is dedicated to George Namisi, Rev. George Wanjohi, Rev. Hogla, my Dad Josephert Sambuli Cheminon, my brothers and sisters

Contents

Acknowledgments

First, I thank The Almighty God for enabling me, Ms. Anne Wang'ombe, George Namisi, Rev. George Wanjohi, Rev. Hogla, Dr. Harrison Kalambuka Angeyo, my Dad Josephert Sambuli Cheminon, my brothers and sisters and all those who supported, encouraged and prayed for me to make it successful . . . .

Dominic Ingosi Sambuli

Nairobi, 2017.

KEYWORDS

Endemic zone———–areas prone to infections

Odd————it is the ratio of chances of occurrence of an event against chance of the same event not occurring

Odd Ratio———This is the comparison of odd in one group against odd of another group.

Wealth quintile———–five groups (in intervals of 20 percent) in which different households are categorized based on the amount of wealth each owns.

Ceterisparibus——— other factors kept constant

Response variable——— output variables dependent on predictors in a model

Predictor variable——- variables that affect or determine changes in response variable

Error term———–factors not captured in data used in analysis variables yet affect output

Intercept———–the point of intersection between y-axis and the curve

Partial regression coefficient——–coefficient which shows the change on the response variable for a unit change in one predictor variable while keeping other predictors constant

Risk factors———–used to refer to predictors in medical researches

Dummy variable———–binary variable with all cases that falls in a certain category takes a value 1 and all cases not in the category take the value zero.

Saturated model——- it is a perfect fit model with as many parameters as data values.

Null model —— model with intercept only and has no predictors

Response/output variables—- dependent factors whose output are obtained after carrying out the analysis.

Parameters—— coefficients of regressions used to give the prediction

Models—— formulas used to reveal relationship between predictors and response variables

odd——-proportions of probability of an event occurring against the probability of the same event not occurring. They are computed by the formula below.

Odd Ratio—— a measure of odd of event occurring in group A compared to the odd of the similar event occurring in group B. they are critical in comparing relative odds of event of interest occurring

Abbreviations

aver. wea—-Average Wealth Quintile GoK——Government of Kenya

KNBS—–Kenya National Bureau of Statistics

NGOs—–Non-Governmental Organizations

WHO—–World Health Organization

S.U.N—–Slept Under Net

LR—-low risk

L—-lake basin

C—-coastal

S—-semi-arid

1 INTRODUCTION

There are various diseases that affect existence of mankind. Some are curable; others are preventable while others are beyond human control. They are spread in various ways and thrive in different environments. Malaria is one of them and has shaken resources of economies globally for decades. Researchers have studied the disease in different ways and have come up with findings about malaria.

Statistically, it has been noted that the most affected people in the society are women and children age 6 months to 14 years. High cases of infection are being recorded in Lake Basins and coastal regions of countries in the tropics of cancer. Donors and governments use these findings to allocate resources towards protection and treatment of malaria.

Resources towards treatment and prevention of malaria are injected mostly in Malaria prone regions. However, the statistical records they use may no longer be viable since factors influencing spread of malaria have been changing for decades. It is thus essential to evaluate whether it is enough for any given organization to focus in Malaria endemic areas only,expecting to gain full control of disease infection.

This study aim at investigating whether governments, non-governmental organizations and private firms will achieve millennium development goals of eradicating malaria infection and its spread by just concentrating in WHO endemic zones by directing most of the funds to women in the age bracket at the risk of the children who are also prone to the disease.

For decades, malaria has been one of the epidemics that threaten human existence on the surface of the earth. By the year 2013, approximately 627,000 people died of Malaria worldwide, large percentage being African children of less than 5 years. This is according to the journal written by Snow et. al [Snow15]. They used Bayesian conditional Auto-Regressive Generalized Linear Mixed Model to analyze malaria trends by studying rainfall amount, resistance to drugs and nets treated with insecticides as predictors. In their journal, it is evident that more curative and preventive measures against malaria are given to adults than children in Kenya and most parts of Africa [Snow15].

Resources from governments and non-governmental organizations have been directed to control of the disease. In 2013, 2.7 billion US dollars were spent globally just to achieve annual targets for Malaria control as well as elimination that had been set for that year. It has also been discovered that women and children are more prone to malaria infection than men and other adults. This is according to research that was done in Maseno area. In addition, pregnant women are more infected compared to their counterparts. In 2014, out of 35 million women who were pregnant, only 20 million of them received a dose for Malaria as intermittent preventive treatment during pregnancy. On average, 62.5 million children who were suffering from Malaria did not receive Malaria therapy [Jen15].

By 2014, children prevalence to malaria had reduced averagely by 46 percent as compared to 26 percent in the previous year. Infections reduced by 26 percent and there was a decrease in mortality rate by 0.47 globally and 0.54 in African regions classified as prone to malaria [WHO14].
In 2013, the number of death resulting from malaria had reduced to 198 million from 227 million worldwide. In Africa, the decrease in death due to malaria was 43 out of 100 as compared to 25 out of 100 in all the countries where the study was conducted. In getting this percentage, the researchers considered the change in population especially for women and children. Approximately, the population at risk reduced by 30 and 34 percent in every 1000 worldwide and in Africa respectively as reported by World Health Organizations. If this rate was to be retained, then incidences of malaria were expected to have decreased by 0.35 worldwide and 0.4 in African regions prone to malaria case by the year 2015 [WHO 2015]. This was a good expectation if the death resulting from Malaria infestations was to be controlled and eradicated from Africa [WHO14].

1.1 Statement of Problems

The fact that a lot of investments and resources are directed to fighting malaria and its significance, the factors leading to spread of the disease are dynamic and thus strategies laid to curb the fever have and may not be very effective. Climate change has led to increase in temperature in areas that were initially too cold for malaria vectors to thrive[David11]. This has created favorable environment for breeding of anopheles mosquito (responsible for the spread of the disease) even non-endemic zones [KNBS15].

There are cases (and have been increasing every year) of malaria diseases in highland and low risk regions. These areas were initially not prone to the epidemic. Unfortunately, researchers specializing in malaria infection concentrate and put their efforts in Malaria prone regions. The areas are coastal, Lake Basin region of Kenya and other warm temperate regions. International health financiers and governments direct resources for control and elimination of the disease in endemic areas only. They assume that other zones are better off. Indeed they are doing a good job.

1.2 Criticism

It could have been better if the surveys were done randomly across the country, not only in areas spotted by WHO as endemic regions. To achieve the target of complete eradication of Malaria from Kenya, the financiers and governments would also have to allocate resources for control and prevention of Malaria in proportion to demand in various regions. However, they cannot do this unless they statistically get convinced that malaria problem is seriously penetrating in excluded places.

1.3 Objectives

This paper aims at testing statistically evidence in disparity in allocation of resources to control of Malaria. It will also consider the household, gender and level of income. Lastly, the study will also unravel the difference between giving more attention to women than children while controlling the infection and spread of Malaria.

1.4 Application of the findings

The statistical findings will reveal to parties interested in the finding whether it is viable to give most of their attention to women and Malaria endemic areas only, or better if they also consider highlands and other non-endemic areas for complete curbing the problems caused by the disease.

2 LITERATURE REVIEW

2.1 Preliminaries

I acknowledge different authors who have written various material about the disease. In this section, the paper presents research findings done by three different researchers on spread and control of Malaria.

2.1.1 Snow et al.

Changing Malaria Prevalence on the Kenyan Coast since 1974: Climate, Drugs and Vector Control, Public Library of Science.
https://www.ncbi.nlm.nih.gov/pubmed/26107772
Snow et al. (2015) studied malaria prevalence, factors leading to infections and people's behavior pertaining to spread of the disease in coastal part of Kenya. This exercise was implemented with the aim of reducing malaria in Africa in long term unlike in short windows as it had been done by previous researchers. The short term study could give the inferences that would lead to incorrect implications thus inaccurate initiative being implemented [Snow15]. Snow and his colleagues decided to study trend of the disease from 1974 to 2014.
They used models like Bayesian conditional auto-regressive generalized linear mixed model and many others to analyze malaria trends using rainfall amount, resistance to drugs and nets treated with insecticides as predictors. They discovered that: prevalence to falciparum parasite on average remained high for the first 23 years. It then decreased gradually until 2011. It however, started increasing until 2014 when the study ended.
They concluded that decline in Malaria cases was associated with treatment of nets during the period when the decrease was experienced. The rise again in prevalence from 2012 was claimed to be associated with resistance to chloroquine, which was the main cure for Malaria. Moreover, achieving its complete elimination of might not be as predicted due to dynamics that were beyond control of the strategies [Snow15].

2.1.2 Jenk15

Prevalence of Malaria parasites in adults and its determinants in Malaria endemic area of Kisumu County, Kenya Malaria Journal 201514:263
https://link.springer.com/article/10.1186/s12936-015-0781-5

Jenkins et al (2015) discovered that most of researchers had been focusing on malaria prevalence among children but not in adults; who are partly women and prone to Malaria just as children. They decided to carry out random survey on prevalence of Malaria among adults in Maseno area, Kisumu County. Slide microscopy was used to measure the presence of Malaria.
Odd ratios were then utilized in comparing level of prevalence among different groups of area residents. The aim was to determine malaria prevalence and the risks factors associated.
It was found out that, in Maseno Malaria prevalence was 28 percent for adults with female being 50 percent more likely to be infected as compared to men [Jen15].

Conclusion was made that, in Maseno area Malaria prevalence is common among adults. They claimed that the rate was greatly increasing among women. The adults were thus considered to be reservoirs for the fever's spreading parasites, which is then transmitted to children who are under 14 years.

2.1.3 KNBS

Kenya Malaria Indicator Survey 2015 Final Report, vol.1
http://statistics.knbs.or.ke/nada/index.php/catalog/89
KNBS (2015) investigated on Malaria epidemiology in Kenya. They discovered that diversity in its prevalence is largely determined by temperature, altitude and rainfall patterns. Highland areas were considered to be seasonal which varied from one year to another. The areas experience temperature of 18 degrees Celsius and above, highest being recorded during long rains. It is at this time that transmission of Malaria tends to be very high.

Malaria endemic areas include coastal and Lake Victoria basin in western part of Kenya. The regions have favorable environment for breeding of mosquitoes and the spread of plasmodium parasites. In these areas, the life cycle of the vector is very short yet the rate of survival is high. This creates an environment that favors transmission of fever throughout the year [KNBS15].

In the report, Semi-arid lands of Kenya were classified to include northern and South-Eastern parts. Intense transmission of malaria is very short in these regions. They usually have very high temperature and water pools during rainy seasons which act as breeding sites for mosquitoes.
Low risk areas are found in a zone that covers Kenyan central Highlands, Nairobi included. Very low temperature in the areas hinders completion of the disease's parasite life cycle.

However, the report indicates that there has been gradual increase in temperature which leads to changes in hydro-logical cycles that favors breeding of Malaria spreading vectors [KNBS15].

3 METHODOLOGY

3.1 Introduction

This paper will use the analytical tools to reveal relationship between Malaria and other factors. There will be use of Odd Ratios, multiple regression, Logistic Regressions and Log Linear Models. Under methodology, the analytical applications, interpretation and assumption of each model will briefly be explained. R software will mainly be used in analysis and codes be given in the appendix. Other things to be found in appendix will be R-codes and tables which will be displayed in appendices.

3.1.1 Multiple Regression Model

It is a technique that enables additional variables to be used in analysis of response variable in such a way that, the impacts of each predictor variable on the response variable can be estimated.

3.1.2 Model

Y=$\beta_0+\beta_1 X_1+\beta 2 X_2+...+\beta_k X_k+\mu$
Where, β_0=the intercept, β_k=coefficients of the parameters, Y=the response variable, Xi-various factors that affect response factor (regressors), μ- residuals

3.1.3 Interpretation

y- is the dependent or response variable that relies on predictors

β_0 - is the intercept of A; the point where curve crosses A axis. It gives the value of A when there are no predictor variables.

β_1- are factors not captured in the analysis yet they affect the response variable.

β_2, ... β_k- the coefficients, give the rate at which response variable changes for each unit change on predictor variables.

If it is negative, the response decreases for every unit increase in predictor. It increases for every unit rise in regressors if the coefficient is positive.

3.1.4 Assumptions made

For this model to be effective, it assumes that:

- Expected error terms=0
- Variance of error terms=δ^2
- Covariance($_m$, $_n$)=0 for mn,
- Error term(e) normality (0, δ^2)

3.2 Odds

odd= $\frac{chance \quad of \quad event \quad occurring}{chances \quad of \quad event \quad not \quad occurring}$

3.3 Odds Ratios

This is a measure of odds of event occurring in group A compared to the odds of the similar event occurring in group B. they are critical in comparing relative odds of event of interest occurring. It is given by the formula below.

Odds Ratio= $\frac{odd \quad of \quad event \quad in \quad group \quad A}{odd \quad of \quad event \quad in \quad group \quad B}$

3.3.1 Example

The odds of being infected with Malaria in town V is 0.8382 and that of town W is 1.234 Odd ratio will thus be:

odds ratio= $\frac{odds \quad of \quad town \quad V}{odds \quad of \quad town \quad W} = \frac{0.8382}{1.234}$

3.3.2 Interpretation

Odds ratio has values ranging from zero to positive infinity.

- When the value is less than one, the event of interest is less likely to occur in group on the numerator compared to the group in the denominator.

- In case the odds ratio is one, both groups had equal chances of occurring since they had the same odds.

- Odds ratio greater than one implies that event of interest is more likely to happen in the group on numerator compared to the one on denominator.

3.4 Logistic Regression Model

This is a statistical tool critical in determining relationship between binary categorical variable and a number of predictors. It can either be simple or multiple logistic regressions. For the purpose of this paper, only multiple logistic regression model will be discussed.

3.4.1 Multiple Logistic Regression Model

It is used in testing association between one dependent variable and two or more independent variables at the same time. It takes the form

$$\ln(Yi)=\beta_0+\beta_1X_1+\beta_2X_2 + \dots + \beta_iX_i$$

where
$\beta_1, \beta_2, ..., \beta_i$
are partial regression coefficient
For binary responses, reference level is assigned the value 0 and the response of interest the value 1.

3.4.2 Interpretation

In determining association, statistical significance of the model has to be fitted. Fitted model is then compared with saturated and null model.
Hypothesis is stated in such a way that the null one considers the null model to be a better

fit than the fitted model

Likelihood ratio statistic Is a test used such that, the chi square distribution has degree of freedom equal to number of predictors.

If the p-value is smaller than the significance level, it implies that the model is statistically significant.
Also, in null hypothesis fitted model is assumed to be a better fit than the saturated model.

Deviance statistic
It is used in testing degrees of freedom equals the sample size (number of parameters in the fitted model).
Small p-value implies lack of fit. If results are non significant, then the null hypothesis is true (fitted model is statistically significant).

When interpreting, the following are considered:

- For Positive coefficient, the transformed value of log will be greater than one to imply that event of interest is more likely to occur.

- For negative coefficient, the value of antilogarithm will be less than one to imply that event of interest is less likely to occur.

- Zero coefficients will be transformed to one. Interpretation will be that, the coefficient does not change odds of event in either way.

3.5 Log-Linear

This is a tool used to evaluate association between two or more predictors. A model is fitted to compare expected and observed counts in cells. If there are only two parameters being evaluated, a two-way contingency table is used. The fit is said to be good if the observation and expectation have a difference which is not significant. This is based on deviance statistic or likelihood ratio statistics.

3.5.1 Model

Let α and β be two categorical variables, α has m and β with n categories. The joint distribution for the two variables will be:
P[α_m,β_n]=p_{mn}

m=1,2,3, ... ,i n=1,2,3, ... , j

Marginal distribution of
α and β
will be

P[α_m]=p_m
and
$P[\beta_n] = p_n$

The model will thus be
S$P_{mn} = \mu_{mn}$;
where S is the sample size

3.5.2 Hypothesis Testing

H0: The two variables are independent
H1: There is dependence between the two variables

The tests are thus given as follows

H0: p_{mn} = p_m x p_n

The expected cell count is given by

μ_{mn}=s x $p_m x p_n$

Its natural logarithm will be

$\ln(\mu_{mn}=\ln(s \text{ x } ꝑ_m xpꝑ_n) = \ln(n)+ \ln ꝑ_m+\ln ꝑ_j$

If α has i levels, one of them is taken as reference level and the rest (i-1) are used as dummy variables.
Also, if β has j levels one is used as reference level and j-1 created as dummy variables. In case of interaction, products of dummy variables created [(i-1)x(j-1)] are utilized in obtaining corresponding regression coefficient. Model fit have predictors in the dummy variables together with terms of interaction. This forms a model with i x j parameters.

The two variables will be statistically associated if the saturated model is the best fit.

3.6 Software to be used

There will be use of R-programming and latex in the analysis and printing of the entire document. R is preferred in this paper over other analytic software because of a number of qualities it has. R enables editing of the script thus change in output without straining. It enables the user to design a model relevant for regression. The software is also locally available and free, thus accessible by any reader of the paper who may be interested in testing the codes that will have been used in the analysis. Lastly, R is very suitable for data management. One only needs to store the script of output which can be opened and run any time while anywhere, so long as the computer has the software installed.

This paper is scientific. It will be prudent to prepare it in latex so that it is standard. Latex will enable typing of formulas and computations which may be tedious to do in word. It will also, enable clarity of the document in such a way that, there will be nothing like a table being on two pages. There will also not be jumping of words from line to line. The software will automatically align tables of content, paragraphs and many other things.

3.7 Source of Data

The data to be used in this paper will be obtained from Kenya National Bureau of Statistics (KNBS) website. The data is found in KENADA micro data and the report ‘Kenya Malaria Indicator Survey 2015 Report’[KNBS15].

4 DATA ANALYSIS

4.1 Introduction

Data analysis is the epitome of this paper. In this section, data from various sources will be analyzed in relation to the topic of study. Models and formulas presented in methodology will be applied in analyzing the data. The data to be used are from Kenya National Bureau of Statistics, World Health Organization, Kenya Malaria Indicator Report and other academic and research journals.

4.2 Multiple Regression Model

This model is used in analyzing how Malaria prevalence is affected by a number of factors.

4.2.1 Malaria Prevalence among Children and Mothers

It has been noted from various data collected that, most initiatives are directed towards mothers (especially pregnant ones), protecting them from Malaria while leaving children of age 6 months to 14 years at risk of infection. It is prudent to evaluate whether there may be any difference in infection risks between mothers aged 16 to 49 years and children of age 6 months to 14 years [KNBS15]. Data from three regions (highlands, coastal and Lake Basin) was used to analyze the relationship. After evaluating the data in R at 0.1 significance level, the following was obtained

4.2.2 Model

Absence of Malaria $=27.38 + 45.23 \quad net$

(Participant mother=1, child=0 slept under net=0, did not sleep under net=1)

The p-value for participant and net were 1.00 and 0.0000202 respectively. Null deviance was 6982.1 at 11 degrees of freedom and residual deviance had 843.9 at 9 degrees of freedom.

4.2.3 Interpretation

Since the coefficient of the participant in the model is zero, it implies that malaria infection does not depend on whether the victim is a mother or a child. In fact the p-value (greater than significance level) definitely indicates that the participant as a factor is statistically not significant in the model.
Even so, for every participant that sleeps under net, there is 45.23 percent increase in chances of not being infected with Malaria.

4.2.4 Observation

From journal by Jenkins et al. it was stated that, the most affected people are women. They further emphasized that interested parties to concentrate mostly on are adults who act as reservoirs for Malaria spreading parasites citeJen15. Here, the analysis has shown that infection is regardless of whether the victim is a woman or a child.

4.2.5 Assumption

The analysis was done on assumption that, participants who sleep under nets would not get the disease.

4.3 Odds and Odd Ratios

By using odds and odd ratios to model logistic regressions, this paper will compare demands and impacts of the parameters towards infection in different regions of the country.

4.3.1 Odds

Demand for Nets

In computing odds from the data

- The odd of pregnant women sleeping under net compared to those who did not sleep under net the previous night before data was collected was
  1.604 for highland regions, 3.566 in Lake Basin, 5.135 in coast, 0.736 in semi arid and 0.8832 in low risk areas of Kenya.

This is according to odds computed from 2015 Kenya Malaria Indicator Report.

- The odds of children sleeping under net a night to data collection is as follows: 1.6385 in highlands, 3.049 in the Lake Basins, 3.132 in coastal, 0.9417 in semi-arid areas and 0.8251 In low risk areas

4.3.2 Interpretation

- Pregnant women are 1.604 times more likely to sleep under net in highlands, 3.566 times more likely in Lake Basin, 5.135 times more likely in the coast, 0.736 times less likely in semi arid and 0.8832 times less likely to sleep under nets in low risk areas of Kenya.

- Children are 1.6385 times more likely to sleep under nets in highland areas, 3.046 times more likely in Lake Basins, 3.132 times more likely in coastal, 0.9417 times less likely in semi arid and 0.8251 times less likely in low risk areas of Kenya.

4.3.3 Logistic Regression

Model for Pregnant Women

$\ln y_1$ =2.18L+3.201C+0.459S+0.551LR**Interpretation**

To sleep under net, women are:

2.18 times more likely in Lake Basin, 3.201 times more likely at the Coast, 0.459 times less likely in Semi-Arid and 0.551 times less likely in Low Risk regions as compared to those in Highland Regions.

Model for Children

$\ln y_2 = 1.86L + 1.912C + 0.575S + 0.504LR$

Interpretation

To sleep under net, children are:
1.86 times more likely in Lake Basin, 1.921 times more likely at the Coast, 0.575 times less likely in Semi-Arid and 0.504 times more likely in Low Risk regions as compared to Highland regions of the country.

Observation

The odds of children sleeping under nets is higher in highlands and semi-arid areas as compared to women in the same areas. The use of nets may be due to availability of the nets. However, the differences may be arising from the accessibility to the nets. The interest is therefore to know whether the differences in use of mosquito nets arise from support by government or other sources.

4.4 Sources of Mosquito Nets

Population in regions demarcated by World Health Organization gets support to prevent malaria from government and other donors. The interest for this paper is to test (using odd and logistic regression) whether there is equity distribution of nets by government in all the five regions. This may give a clue about disparity in sleeping under nets by mothers and children in different regions.

4.4.1 Odds by Government

The government had been providing nets to people in different parts of the country before the data was collected. The odd of receiving nets from government by the five regions was as shown below.
0.6051 in highland areas, 0.3928 in Lake Basin, 0.6447 in the coast, 1.045 in semi arid and 0.7668 in low risk areas.
The odd of receiving nets from government tends to be higher in highlands as compared to Lake Basin but lower than that of coast, semi arid and low risk areas.

Interpretation

From the computation, it is evident that the government is 0.6051 times less likely to give nets in highlands areas, 0.3928 times less likely in Lake Basins, 0.6447 times less likely in coastal, 1.045 times more likely in semi arid and 0.7668 less likely in low risk areas of Kenya.
If we compare distribution of nets by government with willingness to use nets among pregnant mothers and children across the country, we can say that the government did not proportionally give nets in ratio in to demand in different areas (gave more where they are least used i.e semi arid and low risk areas) as compared to highland and Lake Basin.

4.4.2 Logistic Regression

The model is used to compare the likelihood of government providing nets in different regions across the country.
lny=$0.6649L + 1.065C + 1.727S + 1.267LR$

Interpretation

The government is 0.6649 times less likely, 0.065 times less likely to provide nets in Lake Basin, 1.727 times more likely to provide nets in Coastal and 1.267 times more likely to provide nets in Low Risk regions as compared to highland regions of the country.

4.5 Odds by Non-Government

From the computations, NGOs net donors are 0.285 times less likely to give nets in highland areas, 0.1614 times less likely in Lake Basin, 0.3004 times less likely in coastal, 0.0277 times less likely in semi arid and 0.0417 times less likely in low risk areas of Kenya.
This is clear that there is disparity in allocation of mosquito nets among the five regions when compared to the demand of the nets.

4.5.1 Logistic Regression

It is also prudent to test the likelihood of non governmental organizations providing nets in the four regions as compared to Highland regions.

4.5.2 Model

lny=$0.5663L + 1.054C + 0.0972S + 0.146LR$

4.5.3 Interpretation

The NGOs are 0.5663 times less likely to provide nets in Lake Basin, 1.054 times more likely to provide nets at the Coast, 0.0972 times less likely in Semi-arid and 0.146 times less likely in Low Risk areas as compared to Highland regions.

Regions	demand in preg. women	demand in children	supply by GOK	supply by NGOs
L	2.18	1.86	0.6649	0.5663
C	3.201	1.912	1.065	1.054
S	0.459	0.575	1.727	0.0972
LR	0.551	0.504	1.267	0.146

4.5.4 Observations

The government allocated more nets in semi-arid and low risk areas than in highland areas for both children and women. This was not as per the demand for the nets between the regions under comparison.
When allocation of nets by government is reviewed, it is evident that the proportion of nets distributed to semi-arid and low risk areas were not in relation to the demand as compared to the need in highland areas.
Non-governmental organizations minimize bias in distribution of nets in proportion to demand. The highest proportion goes to coast which has the highest demand. The second highest proportion goes to Lake Basin regions whose demands are slightly higher than those of Lake Basins.
This implies that there may not be relationship between the use of mosquito nets and the proportion at which the government distribute them across the country. The interest is then to test whether there is any relationship between sleeping under mosquito nets and government's provision. This can be checked using the multiple regression models below.

4.6 Multiple Regression

4.6.1 Relationship between use of the nets and their Sources

Slept under net= 63.563−0.533 government nets + 1.42 NGOs nets

The p-values for government nets and NGOs are o.568 and 0.196 respectively.

Interpretation

- If both the government and NGOs were not providing nets, 63.563 percent would prefer sleeping under nets.

- If government provides nets, sleeping under net would decreases by 0.533 percent for a unit percent increase in nets provided

- There will be increases by 1.42 percent sleeping under net for every unit percent increase in nets provided by NGOs.

This implies that provision of nets by non governmental bodies may encourage sleeping under them. The contrary happens if the nets are from government. The p-values however indicate that both government and NGOs are not statistically significant in defining the sleeping under net at 5 percent significance level.

Assumption

There are no other sources of the nets for all the five regions a part from the two stated. Some of the participants interviewed may have bought their nets from supermarkets yet they were not captured in the model.
The concept of the two predictors not being statistically significant can thus be ignored.

4.7 Comparison of Malaria Cases in Different Zones

To test Malaria prevalence between zones, odd ratios will elaborate vividly. Odd ratio for Malaria in:

- highlands versus Lake Basin =0.364

- highland versus coastal = 0.507

- highland versus semi-arid areas= 1.233

- highland versus low risk areas=1.422
- low risk areas versus semi-arid areas=0.867

4.7.1 Model

By applying logistic regression, the following model will be obtained. lny=0.364L+0.507C+1.233S+0.867LR

4.7.2 Interpretation

- an individual is 0.364 times less likely to get Malaria in highland as compared to Lake Basin
- an individual is 0.507 times less likely to get Malaria highland areas as compared to coastal regions
- an individual is 1.233 times more likely to be infected with the fever in highland as compared to semi-arid areas
- the person is also 1.422 times more likely to be infected in highlands than in low risk areas
- People in low risk areas are 0.867 times less likely to be infected as those in semi-arid areas.

4.7.3 Wealth Quintiles

It is also of interest to determine whether there is relationship between the use of mosquito nets and wealth status. On average, the odd of having wealth is 0.3387 in highland, 0.368 in Lake Basin, 0.1507 at the coast, 0.1521 in semi-arid and 0.2019 in low risk areas. When use of mosquito nets is regressed against wealth quintiles, the model obtained is as shown below

4.7.4 Model

Use of net = 47.1485 + 0.8033 $wealth \quad quintile$

4.7.5 Interpretation

This indicates that there is 0.8033 percent increase in use of net for every unit percent increase in wealth quintile, other factors held constant.

Also, absence of Malaria cases in children is regressed against use of mosquitoes and wealth quintiles. Model obtained is as shown below

4.7.6 Model

No Malaria cases=-43.8204+0.5208use of net+1.2242 wealth quintile

4.7.7 Interpretation

This implies that, number of Malaria cases decreases by 0.5208 percent for every unit increase in use of nets ceteris Paribas.
Also, the case decreases by 1.2242 percent for every percentage increase in wealth status other factors kept constant.

4.8 LogLinear Regression

4.8.1 Association between Malaria Cases and Regions

From previous analysis, allocation of resources for Malaria control was not proportional to demand for some regions; especially by government. To conclude and recommend about strategies aimed at controlling Malaria, it is essential to test association between the two variables. This model enables testing whether there is any association between the two factors (malaria cases and regions). When the data was analyzed in R, the following output was obtained.

4.8.2 Output

variable	estimate	std error	z-value	pr(>|z|)
intercept	-161.351	297086.518	-0.001	1
m	3.259	5930.997	0.001	1

Null deviance: 1.3863e+01 on 9 degrees of freedom

Residual deviance: 4.2043e-10 on 8 degrees of freedom
AIC: 4

4.8.3 Model

lny=-161.351+3.259m

4.8.4 Interpretation

p-value is larger than the α=0.05. This implies that the fitted model is a good fit
Deviance statistics is very small (4.2043* 10^-10).
$This\ implies\ that\ there\ is\ statistical\ association\ between\ Malaria\ cases\ and\ regions(zones)$

5 CONCLUSION RECOMMENDATION

5.1 Conclusion

This paper has explained the five Malaria endemic zones as classified by World Health Organizations. A part from basic analytical tools used by researchers and analysts, in this paper logistic regression models have been utilized to show disparities on how resources, mostly nets, are allocated towards prevention and treatment of Malaria in different parts of the country.

The proportions at which people need to deal with cases of the fever differ from one region to another. Also, the likelihood of its infection is not the same across the country. It is unfortunate that the government of Kenya rarely considers the fact that not at all will the areas perceived to be non-endemic zones be safe. In fact, some of the areas need more preventive measures than the zones marked to be endemic.

For instance, highlands and low risk areas have been marked to be regions where Malaria is seasonal or rarely experienced. However, climate changes lead to favorable hydrological cycles that favor completion of life cycle of vector disease. This is why there are cases of Malaria being experienced in mountainous and highland regions of Kenya.

5.2 Recommendation Future Research

It would be better for Kenyan government not to discriminate these areas while allocating resources for prevention and treatment of Malaria. To eliminate the disease completely as one of the millennium development goals of the country, it will be critical for eradication strategies to be laid in highland, mountainous and low risk areas.

Bibliography

[David] David I. Stern, Peter W. Gething, Caroline W. Kabaria, William H. Temperley, Abdisalan M. Noor, Emelda A. Okiro, G. Dennis Shanks, Robert W. Snow and Simon I. Hay *Temperature and Malaria in Highland East Africa: Monthly malaria cases at Kericho Unilever Tea Kenya Ltd Hospital*
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0024524

[Feu16] *ManualforPackage PgfplotsTable Component of pgfplots, Version 1.13*
https://sourceforge.net/projects/pgfplots/

[Jen15] Jenkins R., Omollo R., Ongecha M.,Sifuna P.,Othieno C.,Ongeri L., Kingora J. and Ogutu B.*Prevalence of malaria parasites in adults and its determinants in malaria endemic area of Kisumu County, Kenya Malaria Journal201514:263*
https://link.springer.com/article/10.1186/s12936-015-0781-5

[KNBS11] http://statistics.knbs.or.ke/nada/index.php/catalog/10

[KNBS15] KNBS*Kenya Malaria Indicator Survey 2015 Final Report, vol.1*
http://statistics.knbs.or.ke/nada/index.php/catalog/89

[KNBSGok 15] *Survey Methodology: Kenya Demographic and Health Survey.*
http://statistics.knbs.or.ke/nada/index.php/catalog/74

[KNB15] KNBS, NACC, KEMRI,NCPD and Ministry of Health *Kenya Demographic and Health Survey 2014. Key Indicators, pg40-48*
http://statistics.knbs.or.ke/nada/index.php/catalog/74

[Man07] Manning Logistic Regression (with R)
file:///F:/logistic.pdf

[Snow15] Snow R. W., Kibuchi E., Karuri S. W., Sang G., Gitonga C. W., Mwandawiro C., Bejon P. and Noor A. M. Changing Malaria Prevalence on the Kenyan Coast since 1974: Climate, Drugs and Vector Control, Public Library of Science.
https://www.ncbi.nlm.nih.gov/pubmed/26107772

[WHO15] WHO, K.N.B.S and Ministry of Health. *Kenya Stepwise Survey FOR NON COMMUNICABLE DISEASES RISK FACTORS 2015 REPORT*
http://statistics.knbs.or.ke/nada/index.php/catalog/90

[WHO14] WHO14*World Health Organization Report in Kenya. Pg.65.* https://books.google.co.ke/books?id=rg4LDgAAQBAJpg=PA65lpg=PA65dq=world+health+orga
$ensa = Xredir_esc = yv = onepageq = world$

[WHOH15] WHO Health Report *World Health Organization Report. World Malaria Report: Regional and Continental.*

6 APPENDICES

6.1 Appendix 1

6.1.1 R-codes

Simple and multiple regressions models

```
Fit=lm(number of malaria cases temperature)
Fit1=lm(slept under net  government nets + NGOs net)
Fit2=lm(net use wealth)
Fit3=lm(malaria cases use of net+wealth quintile)
```

Multiple Logistic Regression Models

```
m=c(0,1,0,0,0,1,1,1,1,0)
w=c(25.3,26.9,13.1,13.2,16.8,25.3,26.9,13.1,13.2,16.8)
u=c(0,0,0,1,1,0,0,0,1,1)
uw=c(w*u)
p=c(1,1,1,1,1,0,0,0,0,0)
r=c(36.1,53.4,39.3,32.4,27,45.4,77.4,75.8,40.2,38.2)
pw=c(p*w)
pr=c(p*r)
pu=c(p*u)
fit=glm(m w+u+p+r+uw+pw+pr+pu, family=binomial,data=)
summary(fit)
```

6.1.2 Loglinear Regression Model

```
m=c(40.75,59.25,65.4,34.6,56.55,42.45,35.8,64.2,32.6,67.4)
r=c(0,1,1,0,1,0,0,1,0,1) fit1=glm(r m,family=binomial,data=)
summary(fit1)
```

6.2 Appendix 2

6.2.1 tables

Table 1:Multiple regression models (malaria prevalence among children and mothers) Data in percentage obtained from KNBS (Kenya Malaria Indicator Survey, 2015) [KMIS]page 53 and 55 for children and women respectively. Those percentage of those who did not sleep under net was obtained after deduction of 100 percent by less by the percentage who slept under any net.

PARTICIPANTS	WHERE SLEPT	WHERE SLEPT	WHERE SLEPT	
Mothers	undernet	61.6	83.7	78.1
mothers	not under net	38.4	16.3	21.9
children	under net	62.1	75.8	75.3
children	not under net	38.8	24.2	24.7

Table 2: The data in percentage of women and children (used to calculate Odds, Odd Ratios and logistic regression model) from different places of Kenya as from Kenya Malaria Indicator Survey 2015 Final Report [KMIS]. For those who slept under net (table 3.6 on pages 53 table 3.7 on 55 for children women respectively, for average wealth quintile was computed from table 2.5 page 37,Malaria fever from table 4.2 page 66 for women and table 5.1 on page 73 for children) (KNBS). The average malaria fever cases for the women and children was then computed

s.u.n		govt	others	aver. wea	Malaria Cases			region
women	kids	GOK	Others	aver. wea	kids	women	average	region
61.6	61.2	37.7	22.2	25.3	36.1	45.4	40.75	highland
78.1	75.3	28.2	13.1	26.9	53.6	77.4	65.4	lake basin
83.7	75.8	39.2	23.1	13.1	39.3	75.8	57.55	coastal
42.4	48.5	51.1	2.7	13.2	31.4	40.2	35.8	semi-arid
46.9	43.4	29.8	4	16.4	27.0	38.2	32.6	low risk

Table 3: Table below show average in percentage of malaria cases among women and children as obtained from the table above (table 2). The average percent of participants who did not have fever was obtained by 100 percent less the average percent for fever.

This data was used in analysis using Loglinear regression model.

region	code	Marital Status		Total
region	code	positive	negative	Total
highland	no(1)	40.75	59.25	100
lake basin	yes(0)	65.4	34.6	100
coastal	yes(0)	57.55	42.45	100
semi-arid	no(1)	32.6	67.4	100
low risk	no(1)	32.6	67.4	100

6.3 Appendix 3

6.3.1 Maps

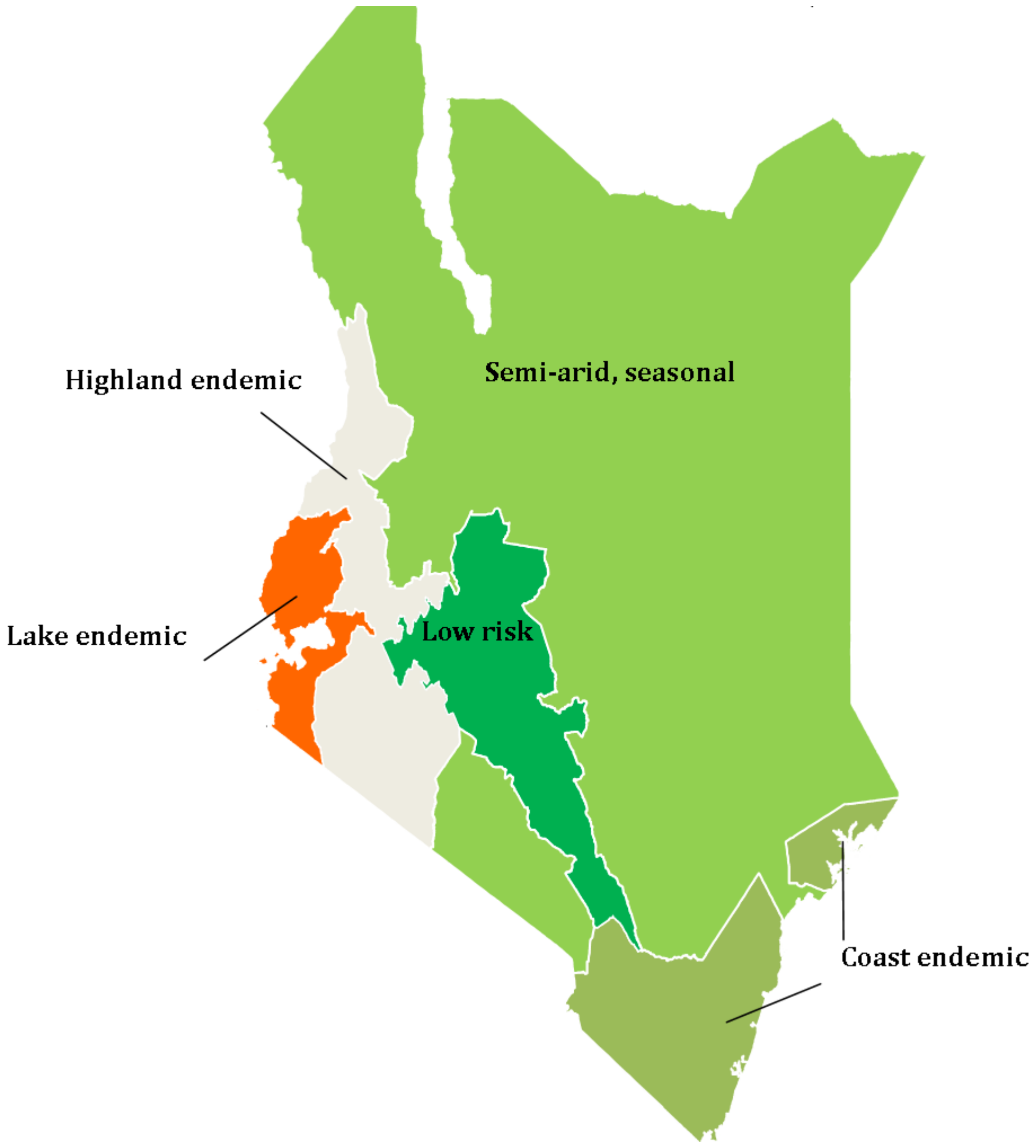

Figure 1. Maps showing the five demarcated zones (Kenya Malaria Indicator Report, 2015