Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 5, Issue, 12, pp.4187-4193, December, 2013

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

INDIVIDUAL KNOWLEDGE AND HOUSEHOLD LEVEL CAUSES ASSOCIATED WITH MALARIA INCIDENCE IN A KUDAPAKKAM VILLAGE: A STATISTICAL EXPLORATION

Chandrasekaran, R., *Manimannan, G. and Prema, R.

Department of Statistics, Madras Christian College, Tambram, Chennai

ARTICLE INFO

Article History:
Received 06th September, 2013
Received in revised form
20th September, 2013
Accepted 08th October, 2013
Published online 25th December, 2013

Key words:

Malaria, Chi-square test,
LLIN,
two step cluster analysis
and Discriminant analysis.

ABSTRACT

A multi-level analysis was carried out among people of Kudapakkam Village. This area was chosen as a model to study malaria Knowledge Attitude and Practices (KAP). This area is situated Thiruvallur district of Tamilnadu in India. The aim is to determine awareness factors associated with malaria. Data were collected from secondary sources containing 580 respondents. The data deals with socio demographic characteristics, knowledge of signs and symptoms, usage of Long Lasting Insecticidal Net (LLIN) and adverse measures of malaria. Forty percent of respondents doesn't know about signs and symptoms of malaria and remaining sixty percentage respondents know the signs and symptoms of malaria. Awareness of malaria, usage of LLIN was significantly associated with family grade of education; Awareness of malaria was not significantly associated with occupation, gender, respondent age, community, occupation of the family. Knowledge, signs and symptoms and usage of LLIN have no association with malaria attacks. In addition, to found the hidden pattern of Knowledge, signs and symptoms and usage of LLIN using two step cluster analysis and cross validate this pattern using discriminant analysis.

INTRODUCTION

Health is the level of functional metabolic efficiency of an organism at both the micro (cellular) and macro (social) level. In the medical field, health is commonly defined as an organism's ability to efficiently respond to challenges. An another widely accepted definition of health is that of the World Health Organization (WHO) which states that Health is a state of complete physical, mental and social well being and not merely the absence of disease on infirmity. Malaria is an infectious disease caused by a parasite (plasmodium) which is transmitted from human to human by the bite of infected female Anopheles mosquitoes. Malaria is one of the most successful parasites ever known to mankind. After thousands of years, it remains the world's most pervasive infection, affecting at least 91 different countries and 300 million people. The disease causes fever, shivering, joint pain, headache, and vomiting. In severe cases, patients can have jaundice, kidney failure, anemia, and can lapse into a coma. It is ever-present in the tropics and countries in sub-Saharan Africa, which account for nearly 90 percent of all malaria cases. The majority of the remaining cases are clustered in India, Brazil, Afghanistan, Sri Lanka, Thailand, Indonesia, Vietnam, Cambodia, and China. Malaria causes 1 to 1.5 million deaths each year, and Africa, it accounts for 25 percent of all deaths of children under the age of five.

Corresponding author: Manimannan, G.
Department of Statistics, Madras Christian College, Tambram, Chennai

Review of Literature

Prevention of the disease through better knowledge and awareness is the appropriate way to keep it away. And people remain healthy, as illness confusion and health-seeking behaviour may enhance or interfere with the effectiveness of control measures (Klein *et al*., 1998). Studies pertaining to knowledge, attitude and practices (KAP) showed that direct interaction with community, plays an important role in circumventing malaria problem (Collins *et al*., 1997; Singh *et al*., 1998). Poor people are at increased risk both of becoming infected with malaria and of having this more frequently. Child mortality rates are known to be higher in poorer households and malaria is responsible for a substantial proportion of their deaths. Poor families live in dwellings that offer little protection against mosquitoes and are less able to afford possessing insecticide-treated nets. Poor people are also less likely to be able to pay either for effective malaria treatment or for transportation to a health facility capable of treating the disease (WHO, RBM, 2003). To analyze the Knowledge, Attitude and Practice (KAP) in malaria control of the respondents using frequencies and chi-square test. In this paper attempt to identify the awareness based on age, education and sex using TwoStep cluster analysis. In addition, to cross-validate the TwoStep cluster analysis using Discriminant Analysis (DA).

Database and Methodology

In this study, the secondary data containing 580 respondents were selected and the information was based on their

respondents Knowledge, Attitude and Practice (KAP) of malaria control in Kudapakkam village, Thiruvallur District. Kudapakkam village consists of two blocks. In 2010, it is noted that people from one block were mostly affected by malaria. Data were collected from health workers using simple random sampling method. The following flow charts describing variables.

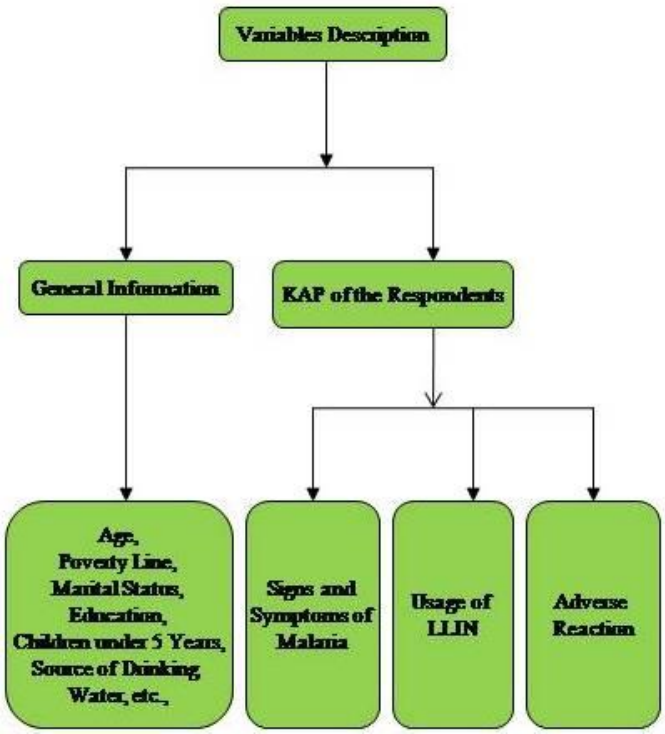


Fig. Variable Description

Algorithm for Two Step Cluster Analysis

The Two Step Cluster Analysis method is an exploratory tool designed to reveal natural groupings (or clusters) within a data set that would otherwise not be clear. The algorithm employed by this procedure has several desirable features that differentiate it from traditional statistical clustering techniques:

Step1: The ability to create clusters based on both categorical and continuous variables.
Step 2: Automatic selection of the number of clusters.
Step 3: The ability to analyze huge data files efficiently.

Clustering Principles

In order to handle categorical and continuous variables, the Two Step Cluster Analysis procedure uses a likelihood distance measure which assumes that variables in the cluster model are independent. Further, each continuous variable is assumed to have a normal (Gaussian) distribution and each categorical variable is assumed to have a multinomial distribution. Empirical internal testing indicates that the procedure is fairly robust to violations of both the assumption of independence and the distributional assumptions, but you should try to be aware of how well these assumptions are met. The two steps of the Two Step Cluster Analysis procedure's algorithm can be summarized as follows:

Step 1. The procedure begins with the construction of a Cluster Features (CF) Tree. The tree begins by placing the first case at the root of the tree in a leaf node that contains variable information about that case. Each successive case is then added to an existing node or forms a new node, based upon its similarity to existing nodes and using the distance measure as the similarity criterion. A node that contains multiple cases contains a summary of variable information about those cases. Thus, the CF tree provides a capsule summary of the data file.
Step 2. The leaf nodes of the CF tree are then grouped using an agglomerative clustering algorithm. The agglomerative clustering can be used to produce a range of solutions. To determine which number of clusters is "best", each of these cluster solutions is compared using Schwarz's Bayesian Criterion (BIC) or the Akaike Information Criterion (AIC) as the clustering criterion (BIRCH by Zhang *et al*, 1996)

Number of clusters: auto-cluster

A characteristic of hierarchical clustering is that it produces a sequence of partitions in one run: 1, 2, 3,…, clusters. A K-means algorithm would need to run multiple times (one for each specified number of clusters) in order to generate the sequence. To determine the number of clusters automatically, SPSS developed a two-step procedure that works well with the hierarchical clustering method. In the first step, the BIC or AIC for each number of clusters within a specified range is calculated and used to find the initial estimate for the number of clusters. In the second step, the initial estimate is refined by finding the largest increase in distance between the two closest clusters in each hierarchical clustering stage.

The BIC and AIC for J clusters are defined as

$$BIC(J) = -2\sum_{j=1}^{J} \P_j + m_j \log(N)$$

$$AIC(J) = -2\sum_{j=1}^{J} \P_j + 2m_j$$

Where

$$m_j = J\left\{2k^A + \sum_{k=n}^{K} L_k - 1\right\}$$

Distance Measure

A distance measure is needed in both the pre-cluster and cluster steps. Two distance measures are available.

Log-Likelihood Distance

The log-likelihood distance measure can handle both continuous and categorical variables. It is a probability based distance. The distance between two clusters is related to the decrease in log-likelihood as they are combined into one cluster. In calculating log-likelihood, normal distributions for continuous variables and multinomial distributions for categorical variables are assumed. It is also assumed that the variables are independent of each other, and so are the cases. The distance between clusters j and s is defined as

$$d(j,s) = \P_j + \P_s - \P_{<j,s>}, \text{ where} \quad (1)$$

$$\P_v = \left(\sum_{k=1}^{K^A} \frac{1}{2} \log\left(\pounds_k^2 + \pounds_{vk}^2\right) + \sum_{k=1}^{K^B} \hat{E}_{vk} \right) \quad (2)$$

$$\hat{E}_{vk} = -\sum_{l=1}^{L_A} \frac{N_{vkl}}{N} \log \frac{N_{vkl}}{N} \quad (3)$$

If $\pounds_k^2$ is ignored in equation (2), the distance between clusters *j* and s would be exactly the decrease in log-likelihood when the two clusters are combined. The $\pounds_k^2$ term is added to solve the problem caused by $\pounds_k^2 = 0$ which would result in the natural logarithm being undefined (this would occur, for example, when a cluster only has one case).

Euclidean distance

This distance measure can only be applied if all variables are continuous. The Euclidean distance between two points is clearly defined. The distance between two clusters is here defined by the Euclidean distance between the two cluster centers. A cluster center is defined as the vector of cluster means of each variable. In our research paper, we are using log-likelihood distance measure.

DISCRIMINANT ANALYSIS

RESULT AND DISCUSSION

The *Auto-Clustering* table can be used to assess the optimal number of clusters in our analysis, as shown below.

Table 1. Auto-Clustering

Number of Clusters	Schwarz's Bayesian Criterion (BIC)	Ratio of Distance Measures
1	16089.308	1.353
2	15451.838	1.828
3	15091.936	1.205
4	15088.072	1.321
5	15157.392	1.089
6	15313.423	1.117
7	15491.582	1.096
8	15695.668	1.062
9	15919.301	1.078
10	16154.764	1.027
11	16404.002	1.056
12	16657.945	1.005
13	16921.100	1.102
14	17185.135	1.047
15	17464.191	

The clustering criterion (in this case the BIC) is computed for each potential number of clusters. Smaller values of the BIC indicate better models, and in this situation, the "best" cluster solution has the smallest BIC. However, there are clustering problems in which the BIC will continue to decrease as the number of clusters increases, but the improvement in the cluster solution, as measured by the BIC Change, is not worth the increased complexity of the cluster model, as measured by the number of clusters. In such situations, the changes in BIC and changes in the distance measure are evaluated to determine the "best" cluster solution. A good solution will have a reasonably large Ratio of BIC Changes and a large Ratio of Distance Measures (Table 1).

Table 2. Cluster Distribution

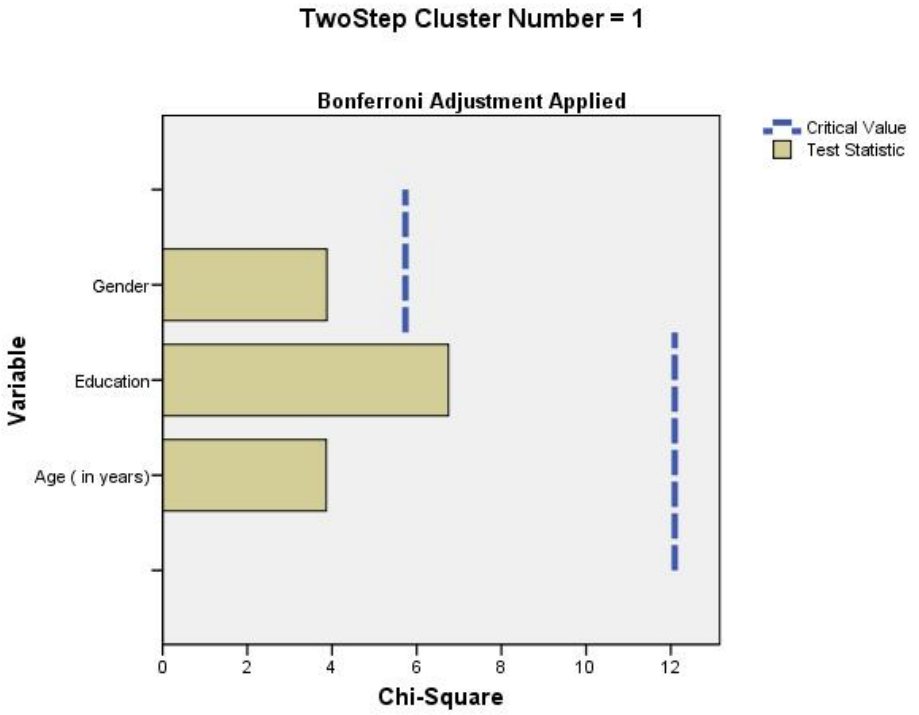
Cluster	Cluster Size	Percentage of Combined	Percentage of Total
1	285	49.1%	49.1%
2	227	39.1%	39.1%
3	68	11.7%	11.7%
Combined	580	100.0%	100.0%

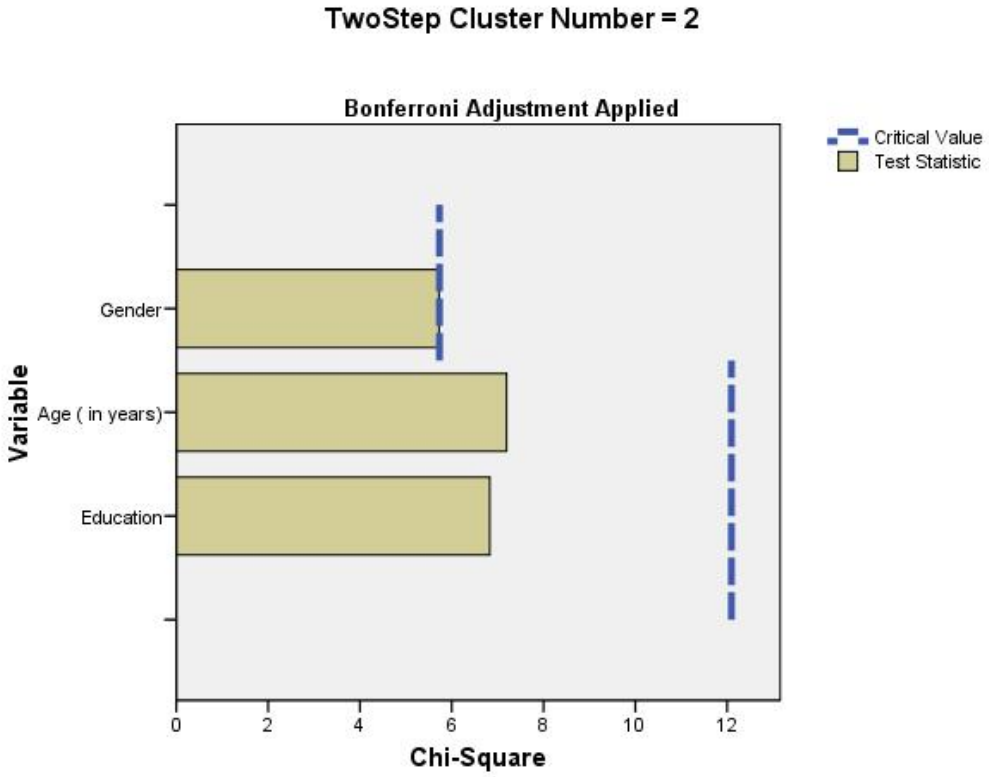
Table 3. Cluster Centroids for KAP

Variable Description	Cluster		
	1	2	3
Drinking Water	1.36	1.47	1.59
How you get LLIN	1.09	1.21	1.12
How does one get malaria?	3.93	4.94	4.53
Do you know signs and symptoms of malaria?	1.31	1.54	1.37
What are the signs and symptoms?	1.91	1.26	1.68
Where does mosquito breed?	2.27	2.51	2.37
Do you think malaria is a serious problem?	1.02	1.30	1.21
Do you know that occurrence of malaria can be reduced?	1.07	1.40	1.32
How occurrence of malaria can be reduced?	1.24	.92	1.00
When you experienced mosquito nuisance more?	1.61	1.38	1.57
What do you actually do to reduce mosquito nuisance?	3.14	3.31	3.40
Aside from LLIN bed net given, do you have any other bed-net?	1.55	1.70	1.63
How many bed net you have	2.81	2.52	2.72
Did you use LLIN bed net previous night?	1.04	1.50	1.24
How many LLIN used last night?	1.65	.77	1.44
Did anybody use other net ?	1.72	1.85	1.74
How many days in a week you used LLIN during sleep?	6.66	4.26	5.63
Do you think bed net is convenient to use?	1.10	1.24	1.16
Is there any pregnant woman in your household?	2.00	1.98	1.03
Did she sleep inside LLIN mosquito net the previous night?	2.00	2.00	1.04
How frequent does she sleeps inside LLIN net?	.00	.00	1.18
Do you think malaria can affect pregnant mother?	1.14	1.34	1.21
Do you experienced any adverse reaction using LLIN	1.99	1.77	1.84
What adverse reaction you experienced?	.00	.51	.41
Did anybody suffered from fever during past two weeks?	1.85	1.70	1.74
Where you go for treatment for fever cases?	1.24	1.62	1.29
Have you seen/heard any advertisement on malaria?	1.31	1.55	1.34
How you come to know about malaria prevention?	2.47	1.74	2.53
Do you think health messages are necessary for malaria prevention?	1.02	1.21	1.13

The cluster distribution Table 2 shows the frequency of each cluster. Of the 580 total cases, were included to the analysis and assigned to clusters, 285 were assigned to the first cluster, 227 to the second, and 68 to the third cluster respectively. The centroids show that the clusters are well separated by the continuous and categorical variables (Table 3). KAP in cluster 1 about Malaria, LLIN and Signs and symptoms are considerably very Low. In Cluster 3 KAP is high and cluster 2 KAP is Moderate level.

Cluster 1's *gender, age and education* variables were important in differentiating this cluster from the other clusters. Cluster 2's *age, education an*d gender variables were important in differentiated this cluster from the other clusters. Cluster 3's *education, age* and gender variables were important in differentiated this cluster from the other clusters.





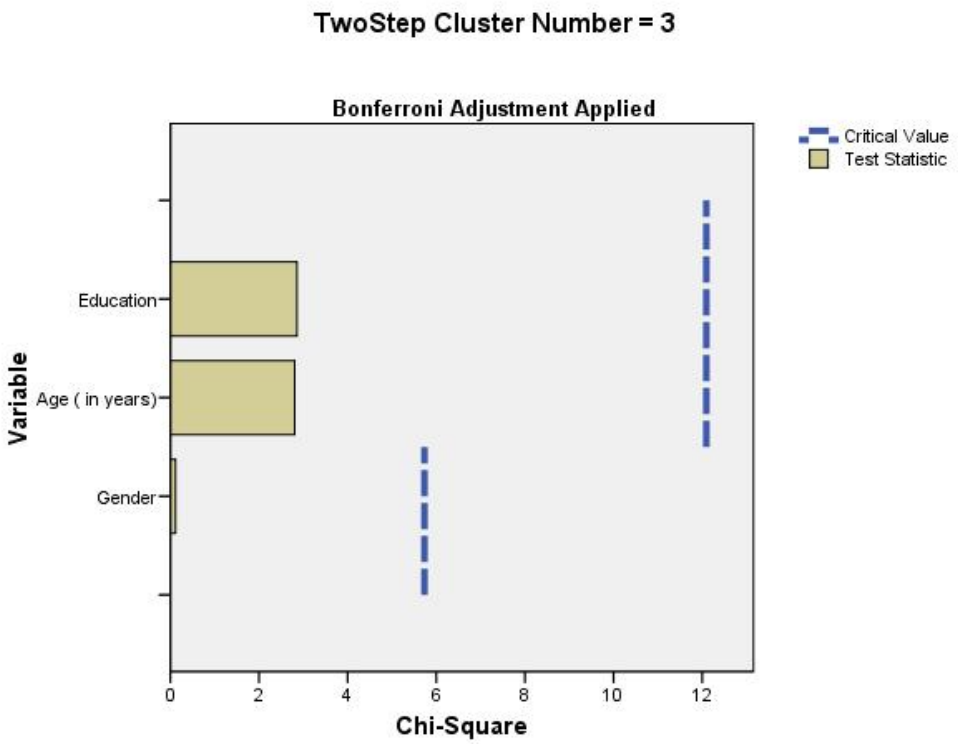
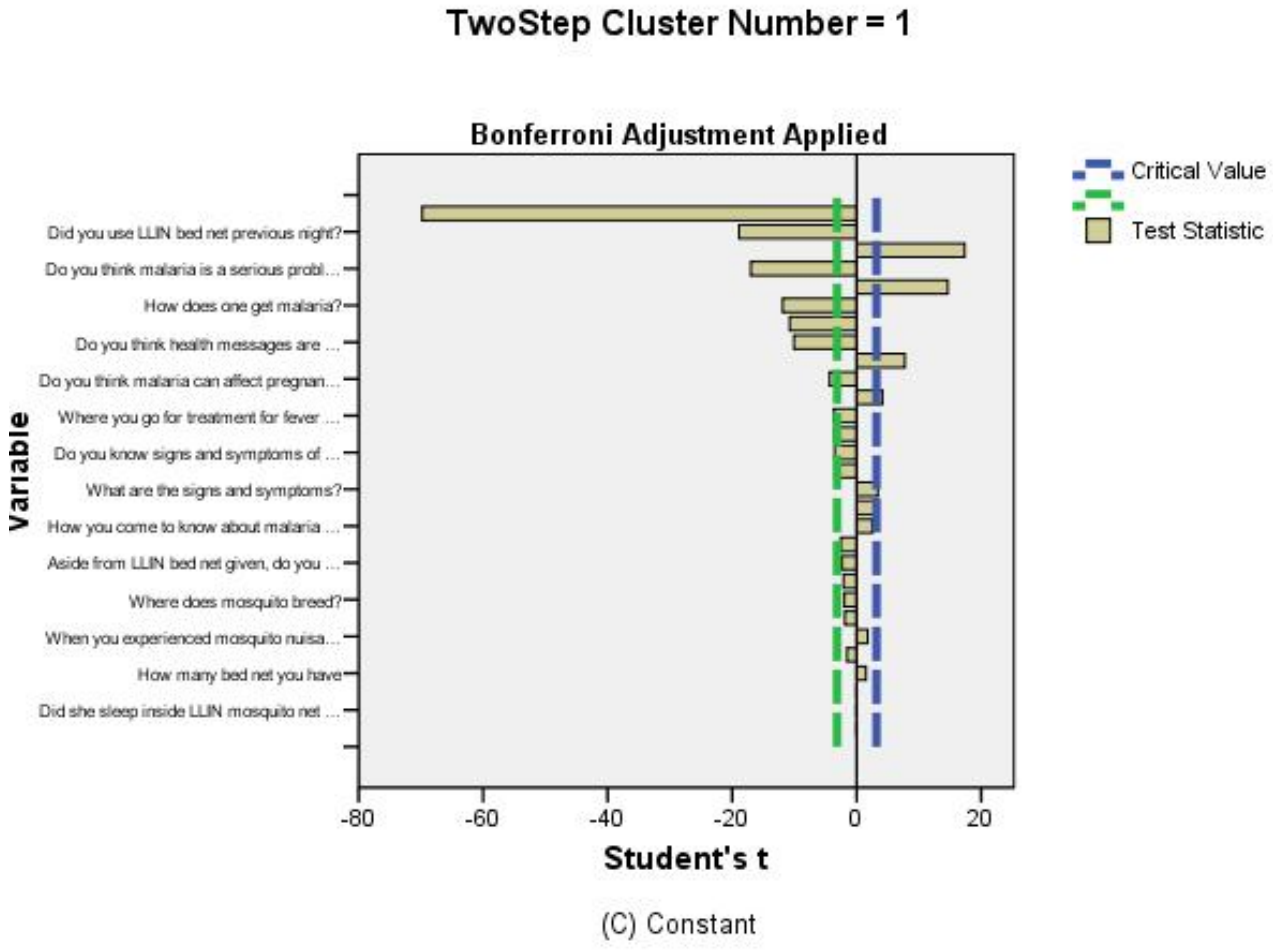
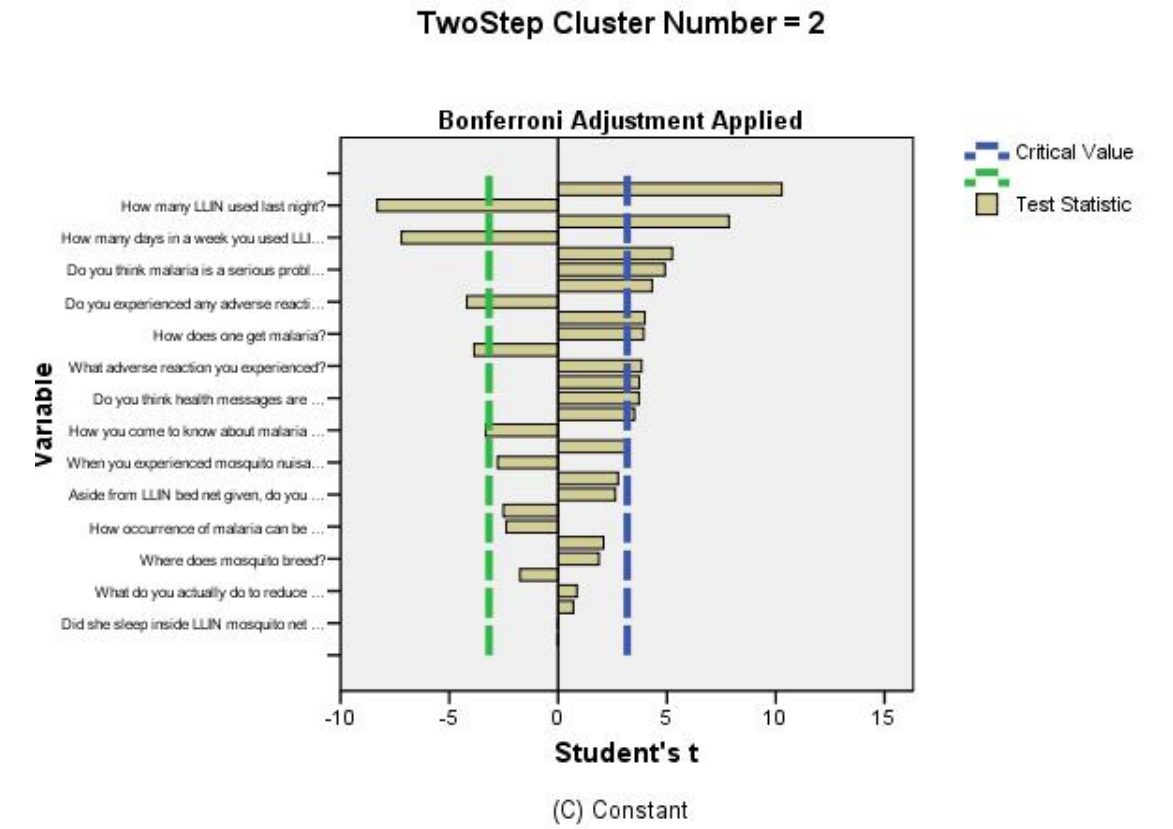
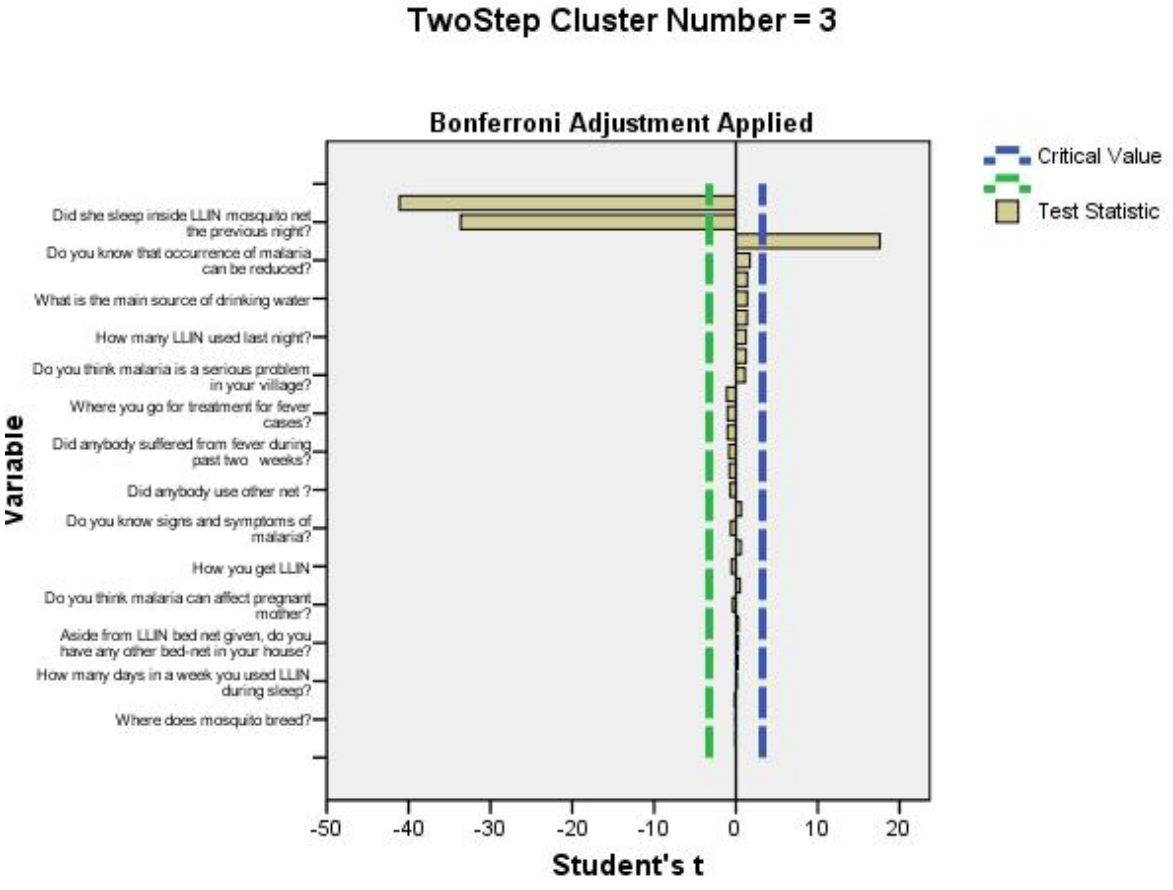


Figure 1 and 2. BP Systolic and Diastolic Cluster







Figures 3. Categorical variable and their Importance

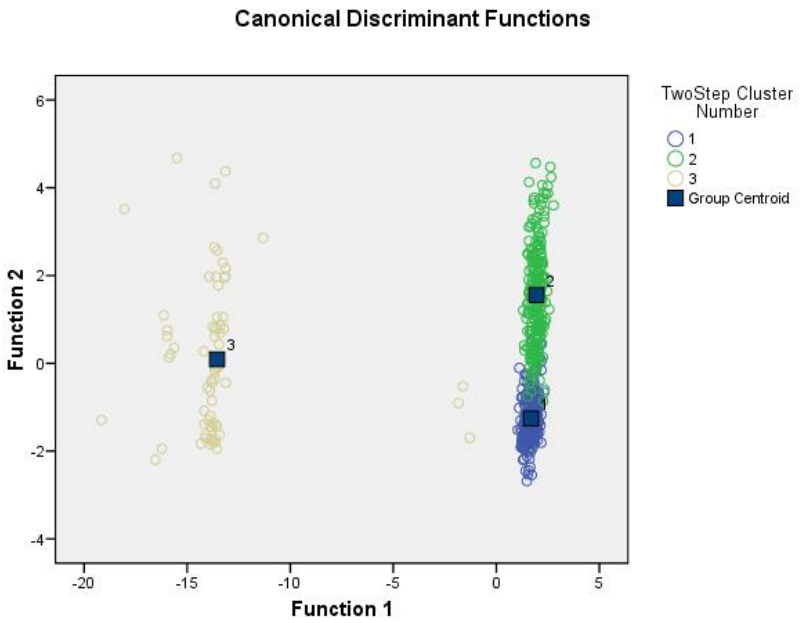


Table 4. Classification table

TwoStep Cluster Number	Predicted Group Membership			
	1	2	3	Total
1	278	7	0	285
2	21	206	0	227
3	3	0	65	68
94.7% of original grouped cases correctly classified.				

The continuous variable wise importance plots are printed for each cluster. On the X-axis is the Student's *t* statistic and on the Y-axis is the continuous variable. If bar exceeds the critical value line, then it indicates that the variable is important in distinguishing the clusters from each other. The plots below show that Clusters 2 are differentiated by mean KAP of the respondents in both (negative and positive) direction, while Clusters 1 and 3 is differentiated by KAP in a positive direction. The classification table 4 shows the practical results of using the discriminant model. Of the cases used to create the model in first cluster, 278 of the 285 respondents who previously defaulted are classified correctly. In second cluster, 206 of the 227 respondents who previously defaulted are classified correctly. In third cluster, 65 of the 68 respondents who previously defaulted are classified correctly. Overall, 94.7% of the cases are classified correctly. Classifications based upon the cases used to create the model tend to be too "optimistic" in the sense that their classification rate is inflated.

Conclusion

From the result of chi-square and t-test, we reject the null hypothesis and conclude that the variables gender, education and age of KAP level of malaria is not the same. The centroids show that the clusters are well separated by the continuous and categorical variables. KAP in cluster 1 about Malaria, LLIN and signs and symptoms are considerably very Low. In Cluster 3 KAP is High and in cluster 2 KAP is Moderate using Two Step cluster analysis. Cross-validation achieved 95 percent of the original groups were classified correctly and rest of them misclassified using Discriminant analysis. This case study is useful for a malaria KAP care which intends to split the respondents, for a better KAP. When new patients turn up for KAP proper guidance can be given based on TwoStep and Discriminant Classification.

Acknowledgement

We would like to express our grateful thanks to D. Aswini Kumar Maji, Health officer for Providing us the data for research.

REFERENCES

Collins KA, Samuel KD, Edwin A A, Kwadwo A, Korum, Francis KN. Malaria related beliefs and behaviour in southern Ghana: implications for treatment, prevention and control. Trop Med Int Hlth 1997; 2(5): 488–99.

Klein RE, Weller SC, Zeissing R, Richards FO, Ruebush TK. Knowledge, belief and practices in relation to malaria transmission and vector control in Guatemala. Am J Trop Med Hyg. 1995; 52: 383–8.

Manimannan G., S. Hari and G. Vijay Thiraviyam (2013), Data Mining Applications in Master Health Checkup: a Statistical Exploration, International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 2.

Norusis, M. 2004. *SPSS 13.0 Statistical Procedures Companion*. Upper Saddle-River, N.J.: Prentice Hall, Inc..

Two Step Cluster Analysis, available at http://support.spss.com/productsext/spss/documentation/statistics/algorithms/14.0/twostep_cluster.pdf

****