International Journal of Scientific and Innovative Mathematical Research(IJSIMR)
Volume 8, Issue 10, 2020, PP 1-13
ISSN No. (Online) 2347-3142
DOI: https://doi.org/10.20431/2347-3142.0810001
www.arcjournals.org

The Pattern of Wealth Index and Household Structure of Tamilnadu

A.Poompavai[1], R. Divya[2], R. Lakshmi Priya[3], Manimannan G[4*]

[1] *Assistant Professor, Departemnt of Statistics, Apollo College of Arts and Sacience, Chennai*

[2] *Assistant Professor, Departemnt of Statistics, SDNB Vaishnav College for Women, Chennai*

[3] *Assistant Professor, Departemnt of Statistics, Dr. Ambedkar Govt. Arts College,, Chennai*

[4] *Assistant Professor, Departemnt of Mathematics, TMG College of Arts and Science, Chennai*

Corresponding Author: *Manimannan G, *Assistant Professor, Departemnt of Mathematics, TMG College of Arts and Science, Chennai*

Abstract: *This research paper is attempted to systematically assess the wealth index between the household structure and educational level based on National Family Health Survey-4. The secondary data sources were collected during the period of 2014-2015 National Family Health Survey (NFHS-4) in Tamilnadu, India. To assess of wealth index specifies the economic status of households among the study population. The statistical techniques like, Chi-square and Binary logistic regression model used for the analysis. To identify the statistically significant association were found between Wealth index, Highest educational level and Household structure (p<0.05). In this study the researcher twelve socio-economic variables are considered to compute wealth index. First apply to the Principal Component Analysis (PCA) and extracted factore score of wealth index are named as Lower Class, Middle Class, and Upper Class. In addition, to identify the relationship between household structure and the above class wealth index is examined by using a Binary logistic regression mode* $\boldsymbol{logit}(\boldsymbol{p}) = \boldsymbol{\alpha} + \sum \boldsymbol{\beta_i X_{ij}}$ *Defacto household member and Lower class of wealth index and middle class wealth index were significant predictors for household structure. The overall classification accuracy is 74.4 percentages. The 95% Confidence Interval for odd ratio of defacto household member are [1.817, 1.917], wealth index specified for Lower Class are [0.360, 0.521] and Middle Class are [0.643, 0.820]. The Chi-square statistics establish that the results are rejected the hypothesis between wealth index and household structure, household structure and higher education.*

Keywords: *Chi-square Statistics, Factor Analysis, Binary Logistic Regression and Wealth Index of Household..*

1. Introduction

The National Family Health Survey (NFHS) may be a large-scale, multi-round survey conducted during a stratified sample of households throughout India. Four rounds of the survey are conducted since the primary survey in 1992-93. The survey provides state and national information for India on fertility, infant and child mortality, the practice of birth control, maternal and child health, reproductive health, nutrition, anemia, utilization and quality of health and birth control services.

1.1.1 The primary National Family Health Survey (NFHS-1)

The survey was conducted in 1992-93. The survey collected extensive information on population, health, and nutrition, with a stress on women and young children. Eighteen Population Research Centers (PRCs), located in universities and institutes of national repute, assisted IIPS altogether stages of conducting NFHS-1.

1.1.2 The Second National Family Health Survey (NFHS-2)

The survey was conducted in 1998-99 altogether 26 states of India with added features on the standard of health and birth control services, violence , reproductive health, anemia, the nutrition of girls , and therefore the status of girls .

1.1.3 The Third National Family Health Survey (NFHS-3)

The survey was administered in 2005-2006. Eighteen Research Organizations including five Population Research Centers administered the survey in 29 states of India. The National Family Health Survey (NFHS-3) is that the third during a series of national surveys. All three surveys were conducted under the stewardship of the Ministry of Health and Family Welfare, Government of India, with the International Institute for Population Sciences, Mumbai, serving because the nodal agency. ORC Macro, Calverton, Maryland, USA, provided technical assistance for all three NFHS surveys. NFHS-3 funding was provided by the USA Agency for International Development, the Department for International Development (United Kingdom), the Bill and Melinda Gates Foundation, UNICEF, the United Nations Population Fund, and therefore the Government of India. Assistance for the HIV component of the NFHS-3 survey was provided by the National AIDS Control Organization and therefore the National AIDS Research Institute. In NFHS-3, 18 research organizations conducted interviews with quite 230,000 women age 15-49 and men age 15-54 throughout India. NFHS-3 also tested quite 100,000 women and men for HIV and quite 200,000 adults and young children for anemia. Fieldwork for NFHS-3 was conducted from December 2005 to August 2006.

1.1.4 National Family Health Survey (NFHS-4)

In 2014-2015, India will implement the fourth National Family Health Survey (NFHS-4).In addition to the 29 states; NFHS-4 also will include all six union territories for the primary time and can also provide estimates of most indicators at the district level for all 640 districts within the country as per the 2011 census. NFHS-4 sample size is predicted to be approximately 568,200 households, up from about 109,000 households in NFHS-3. this is often expected to yield a complete sample of 625,014 women and 93,065 men eligible for the interview. In these households information on 265,653 children below age 5 are going to be collected within the survey.

NFHS- will provide updates and evidence of trends in key population, health and nutrition indicators, including HIV prevalence. Moreover, the survey will cover a variety of health-related issues, including fertility, infant and child mortality, maternal and child health, perinatal mortality, adolescent reproductive health, high-risk sexual behavior, safe injections, tuberculosis, and malaria, non-communicable diseases, violence , HIV knowledge, and attitudes toward people living with HIV.

2. REVIEW OF LITERATURE

El-Houssainy et.al. (2008) and Diana Le Ngo (2011): A Wealth Index of Households Living Conditions in Mauritania:

Evaluating poverty reduction requires repeated measures of the living standards of the poor. during this paper, the likelihood of constructing the asset index by using data of Mauritanian Survey on Household Living Conditions (SHLC, 2008) are going to be investigated and therefore the relation between household socioeconomic positions classified by using asset index and traditional money metric measures. Household expenditure also will be considered.

Jeroen Smits et. al. (2013) and Seema Vyas and Lilani Kumaranayake ,(2006), : The International Wealth Index (IWI):

This paper was present the International Wealth Index (IWI), the primary strictly comparable asset based index for household's long-term economic status which will be used for all low and middle income countries. IWI is analogous to the widely used wealth indices included within the Demographic and Health Surveys and UNICEF MICS surveys, but adds the property of comparability across place and time. IWI is predicated on data from 2.1 million households in 97 developing countries. With IWI we offer a stable and understandable yardstick for evaluating and comparing the economic situation of households, social groups and societies across all regions of the developing world. A household's ranking on IWI indicates to what extent the household possesses a basic set of assets, valued highly by people all across the world . IWI is tested thoroughly for reliability and validity. National IWI values are highly correlated with the Human

Development Index, anticipation , value and academic outcomes and IWI-based poverty measures are highly correlated with Poverty Headcount Ratios.

Diana Lee Ngo dianakimlee (2012) and S.S Morris**, et. al. (**2000): Measuring Household Wealth Comparably Using an Asset-based Index

Asset indices supported durables ownership are commonly wont to proxy for wealth in surveys lacking detailed income and expenditure data. Yet, the present tools to make these indices are theoretically unfounded, limiting the potential applications for such indices. During this research paper, they discuss the utilization of consumer theory supported lexicographic preferences to motivate the utilization of the index and develop an easy rescaling method to make a wealth index which is comparable across surveys. Using three rounds of LSMS data from Nicaragua, they supply evidence for the method's validity both within and across surveys using expenditure benchmarks. The resulting index performs similarly to the commonly used correlation analysis index within surveys, but it offer some improvement in wealth comparisons across surveys when there are few assets common to all or any surveys.

3. DATA AND METHODOLOGY

The secondary source of database were collected from National Family Health Survey during the period of 2014-2015. In this research paper is to asses of wealth index and their structure in the state of Tanilnadu, India from NFHS-IV.

3.1. Chi-Square Test

The chi-square test is one of the simplest and most widely used non-parametric test in statistical work. The quantity chi-square describes magnitude of the discrepancy between theory (E) and observation(O).There are basically two types of random variables and they yield two types of data, numerical and categorical.

It is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(r-1)(c-1)}$$

Where the square of the differences between the observed (O_i) and expected values (E_i) in each cell, divided by the expected value are added across all of the cells in the table. The distribution of the statistic χ^2 is chi-square with (r-1) (c-1) degrees of freedom, where r represents the number of rows in the two-way table and c represents the number of columns. The chi-square distribution is defined for all positive values in the following figure 1.

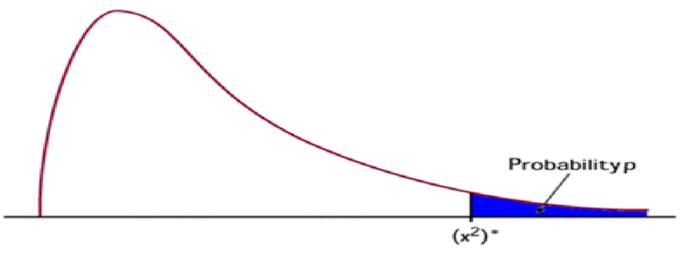

Figure1. *Chi-square distribution Positive values*

The p- value for the chi-square test is P ($\chi^2 \geq x^2$), the probability of observing a value at least as extreme as the test statistic for a chi-square distribution with (r-1) (c-1) degrees of freedom. The strength of association of chi-square is shown below

>0.5 high association

0.3 to 0.5 moderate association

0.1 to 0.3 low association

0.0 to 0.1 little if any association (David E. *et. al.* (2001),

3.2. Factor Analysis

Factor analysis is a data reduction technique used to reduce a large number of variables to a smaller set of underlying factors that summarize the essential information contained in the variables. More frequently; factor analysis is used as an exploratory technique when we wish to summarize the structure of a set of variables (Richard A. and Dean W.1988).

Mathematical model: In matrix notation, the factor analysis model is

$$X_i - \mu_i = l_{ij}F_i + \epsilon_i$$

with the coefficient l_{ij} representing the loading of the i^{th} variable on the j^{th}factor. F represents the vector of unobservable common factors with $F_i$denoting the j^{th}common factor and ξ is the error vector with ξ_i denoting the i^{th} specific factor. A factor analytical procedure comprises of

Computation of the Correlation Matrix: To determine the appropriateness of the factor analytic model.

Factor Extraction: To determine the number of factors necessary to represent the data. A scree plot can also be used to extract the number of factors. The screw plot graphically displays the Eigen values for each factor and the "elbow" showing the predominant factor.

Rotation: To make the factor structure more interpretable. Rotation may be orthogonal (Varimax) or oblique (Quartimax). The choice of rotation is both empirically and theoretically driven. Varimax structure is simple and gives a clear separation of the factors. The method is also proved to be very successful as an analytical approach to obtain an orthogonal rotation of factors\.

3.3. Binary Logistic Regression

Binary logistic regression model is employed to elucidate the connection between the result variable and therefore the factors, by taking the result variable (compliance) to be dichotomous (Compliance and non-compliance).

Logistic regression is additionally called as Logit model. They're wont to model dichotomous outcome variables. Logistic regression measures the connection between a categorical variable and one or more independent variables, which are usually (but not necessarily) continuous, by using probability scores because the predicted values of the variable . The logistic regression model is just a non-linear transformation of the rectilinear regression . The logistic distribution is an S-shaped distribution function (cumulative density function) which is analogous to the quality Gaussian distribution and constrains the estimated probabilities to lie between 0 and 1.

Logistic regression measures the connection between a categorical outcome variable and one or more independent variables, which are usually continuous, by using probability scores because the predicted values of the result variable. Logistic regressions are often binomial or multinomial. Binomial or binary logistic regression deals with situations during which the observed outcome for a variable can have only two possible types like compliance and non-compliance. An evidence of logistic regression begins with an evidence of the logistic function, which always takes on values between zero and one: x

$$\mathbf{F(t) = \frac{e^t}{e^t + 1}}$$

't' is the linear function of an independent or explanatory variable x. The logistic function can be written as

$$\mathbf{F(x)} = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

This will be interpreted as the probability of the outcome variable equaling a "Compliance" rather than non-compliance. Inverse of the logistic is defined as

$$\mathbf{g(x) = \ln \frac{F(x)}{1-F(x)} = \beta_0 + \beta_1 x}$$

And equivalently

$$\frac{\mathbf{F(x)}}{\mathbf{1-F(x)}} = \mathbf{e^{\beta_0} + \beta_1 x}$$

The logistic function is beneficial because it can take an input with any value from negative as infinity to positive infinity, whereas the output is confined to values between 0 and 1 and hence is interpretable as a probability. Within the above equation, refers to the Logit function of some given linear combination of the predictors, denotes the Napierian logarithm , is that the probability that the variable equals a case, is that the intercept from the rectilinear regression equation, is that the parametric statistic multiplied by some value of the predictor, and base denotes the exponential .

A Logistic model is claimed to supply better fit data if it demonstrates an improvement over the intercept only model which is additionally called as “null model”. An intercept only model is good baseline because it doesn’t contain any predictors. Consequently, consistent with the model, all observations would be predicted to belong to its largest outcome category. An improvement over this baseline is examined using the inferential statistics namely likelihood ratio, Wald statistic.

3.4. Univariate logistic Regression

Logistic regression is fitted with the one or experimental variable by using probability scores because the predicted values of the variable . However, Univariate Logistic regression is administered with one experimental variable with the result variable. For every of the independent variables logistic regression is fitted with the result variable.

3.5. Multiple Logistic Regression

The specific sort of the multiple logistic models with k predictors is as follows

$$Logit(p) = \mathbf{b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_kX_k}$$

3.6. Odds Ratio

The odds ratio (OR) is a measure of association which has found wide use, especially in epidemiology, as it approximates how much more likely or unlikely it is for the outcome to be present among those with $x = 1$ than among those with $x = 0$. Hence, for logistic regression with a dichotomous independent variable coded 1 and 0, the relationship between the odds ratio and the regression coefficient is

$$OR = e^{\beta_1}$$

This simple relationship between the coefficient and the odds ratio is the fundamental reason why logistic regression has proven to be such a powerful analytic research tool. The interpretation given for the odds ratio is based on the fact that in many instances it approximates a quantity called the relative risk. The OR is usually the parameter of interest in a logistic regression due to its ease of interpretation. In theory, for large enough sample sizes, the distribution of OR is normal.

In summary, for a dichotomous variable the parameter of interest is the odds ratio. An estimate of this parameter may be obtained from the estimated logistic regression coefficient, regardless of how the variable is coded. This relationship between the logistic regression coefficient and the odds ratio provides the foundation for interpretation of all logistic regression results.

3.7. Calculation of Odds Ratio

Below shown a sample of 2x2 table 1 for simple calculation of odds

Table1. *Contingency table for Chi-square test (2x2).*

	X^-	X^+	**Row Total**
Y^-	a	b	a+b
Y^+	c	D	c+d
Column Total	**a+c**	**b+d**	**a+b+c+d**

The odds ratio is simply the ratio of the two odds,

$$\mathbf{OR} = \frac{\mathbf{a/b}}{\mathbf{c/d}}$$

4. Analysis and Interpretation

4.1. Socio Demographic Variable

Table2. *Frequency table for Demographic variables*

CHARACTERISTICS	CODINGS	FREQUENCY	%
Sex of household member	Male-1	6488	47
	Female-2	7318	53
Household structure	Nuclear-0	5898	42.7
	non-nuclear-1	7908	57.3

The above table 2 shows that the results of 47% respondents are male and 53% respondents are female. In household structure 57.3% are from non-nuclear family and 42.7% from nuclear family.

4.2. Chi – Square Test for Independence of Attributes

Wealth Index* Household Structure

Hypothesis

H_0: There is no association between household structure and wealth index

H_1: There is association between household structure and wealth index

Table 3. *Cross Tabulation for Wealth Index and Household Structure*

		Household structure		
		Nuclear	Non-nuclear	Total
Wealth Index	Poorest	90	52	142
	Poorer	337	219	556
	Middle	794	859	1653
	Richer	2166	3066	5232
	Richest	2511	3712	6223
Total		5898	7908	13806

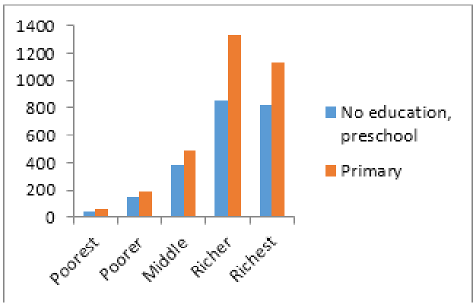

Figure 2. Wealth index of Nuclear and Non- Nuclear Households

Table4. *Chi- square test for Wealth Index and Household Structure*

Chi-Square Tests			
	Value	df	Asymp . Sig. (2-sided)
Pearson Chi-Square	134.589[a]	4	.000
Likelihood Ratio	133.244	4	.000
Linear-by-Linear Association	104.058	1	.000
N of Valid Cases	13806		

Inference

Since the p-value is <0.05 H_0 is rejected. Hence, there is association between household structure and wealth index in Table 3 and 4. The classification is visualized in Figure 2 of household structure and wealth index.

4.3. Household Structure * Highest Education Level Attained

Hypothesis

H_0: There is no association between household structure and highest education level attained.

H_1: There is association between household structure and highest education level attained.

Table5. *Cross Tabulation for Household Structure and Education Level of Nuclear and Non- Nuclear Household*

Count							
		Highest educational level attained					
Household structure		No education, Preschool	Primary	Secondary	Higher	Don't know	Total
	Nuclear	653	1328	3211	690	16	5898
	Non-nuclear	1593	1879	3727	682	27	7908
Total		2246	3207	6938	1372	43	13806

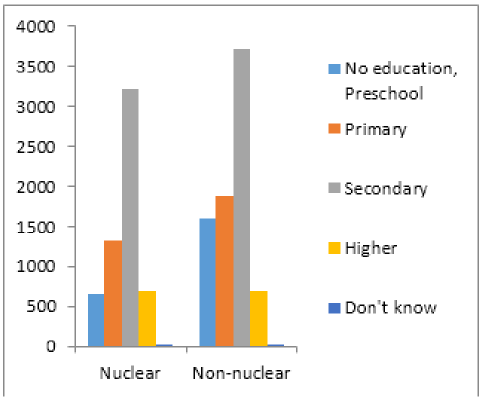

Figure 3. Wealth index of Nuclear and Non- Nuclear Education Level of Nuclear and Non- Nuclear Household

Table6. *Chi- square test for Household Structure and Education Level of Nuclear and Non- Nuclear Household*

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	241.808[a]	4	.000
Likelihood Ratio	248.552	4	.000
Linear-by-Linear Association	182.129	1	.000
N of Valid Cases	13806		

Inference

Since the p-value is <0.05 we rejected H_0. Hence, there is association between household structure and highest education level attained in Table 5 and 6. The classification is visualized in Figure 3 of household structure and wealth index.

4.4 Wealth Index * Highest Education Level Attained

Hypothesis

H_0**:** There is no association between wealth index and highest education level attained

H_1: There is association between wealth index and highest education level attained

Table7. *Cross Tabulation for Education Level and Wealth Index*

Highest educational level attained							
Wealth Index		No education, preschool	Primary	Secondary	Higher	Don't know	Total
	Poorest	46	59	37	0	0	142
	Poorer	144	191	212	6	3	556
	Middle	381	490	750	22	10	1653
	Richer	855	1332	2789	240	16	5232
	Richest	820	1135	3150	1104	14	6223
Total		2246	3207	6938	1372	43	13806

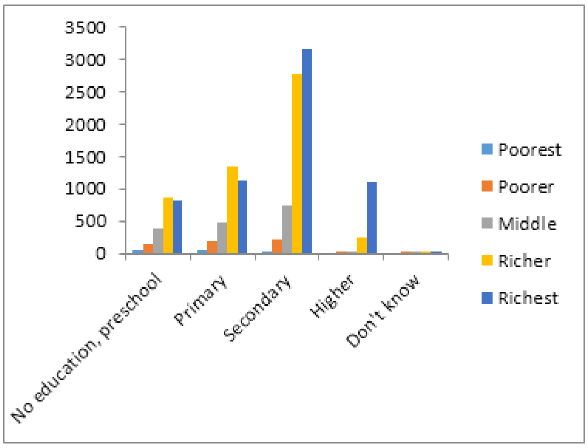

Figure 4. Wealth index of Education Level

Table8. *Chi -square Test for Wealth index of Education Level*

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1066.192	16	.000
Likelihood Ratio	1132.978	16	.000
Linear-by-Linear Association	537.848	1	.000
N of Valid Cases	13806		

Inference

Since the p-value is <0.05 we reject H_0rejected. Hence there is association between wealth index and highest education level attained in Table 7 and 8. . The classification is visualized in Figure 4 of wealth index and education Level.

4.5. FACTOR ANALYSIS

Factor analysis has been performed to identify the factors which assets are combined together. The results are given below. The communalities above, measure the percent of variance in a given variable explained by the entire factor. The "extracted" communalities are the percent of variance in a given variable explained by the factors, which are extracted.

Table9. *The extracted factor Communalities 12 Parameters*

Communalities		
	Initial	Extraction
Number of de facto Household members	1.000	.657
Source of drinking water	1.000	.996
Source of non-drinking water	1.000	.996
Has electricity	1.000	.378
Has television	1.000	.486
Has refrigerator	1.000	.473
Has motorcycle/scooter	1.000	.342
Has car	1.000	.281
Main floor material	1.000	.266
Main wall [exterior] material	1.000	.143
Main roof material	1.000	.117
Household structure	1.000	.667
Extraction Method: Principal Component Analysis.		

The extracted factors explained about in Table 9 attained 99.6% source of drinking water and non-drinking water scores. The minimum factor score explained variable is main roof material (11.7%) in Table 7.. The "Total Variance Explained" table displayed below shows the Eigen values, which are the proportion of total variance in all variables, accounted for by that factor.

Table10. *The Rerated factor matrix and their variance percentage of 12 Parameters*

Componen t	Initial Eigen values			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.419	20.155	20.155	2.315	19.291	19.291
2	1.945	16.208	36.363	2.014	16.787	36.078
3	1.438	11.985	48.347	1.472	12.270	48.347
4	1.346	11.219	59.567			
5	.951	7.928	67.495			
6	.795	6.629	74.124			
7	.735	6.126	80.250			
8	.715	5.955	86.205			
9	.585	4.877	91.082			
10	.569	4.739	95.821			
11	.501	4.179	100.000			
12	.150	2.625	100.000			

From the above table 10, twelve variables are reduced to three factors with 20.15% of the total variation explained by component **1** and 36.363% of the total variation explained by component **2**. The total variation explained in the component 3 with 48.347%.

The initial Eigen values and extraction sum of squared loading columns are the same, except the latter only list factors, which have actually been extracted in the solution. The “Rotated Sum of Squared Loadings” gives the Eigen values after rotation improves the interpretability of the factor with Varimax rotation. The total variance explained is the same but rotation changes in the Eigen vales for each of the extracted factors. The Screw Plot Figure 4 also gives an idea about the number of factors to be extracted.

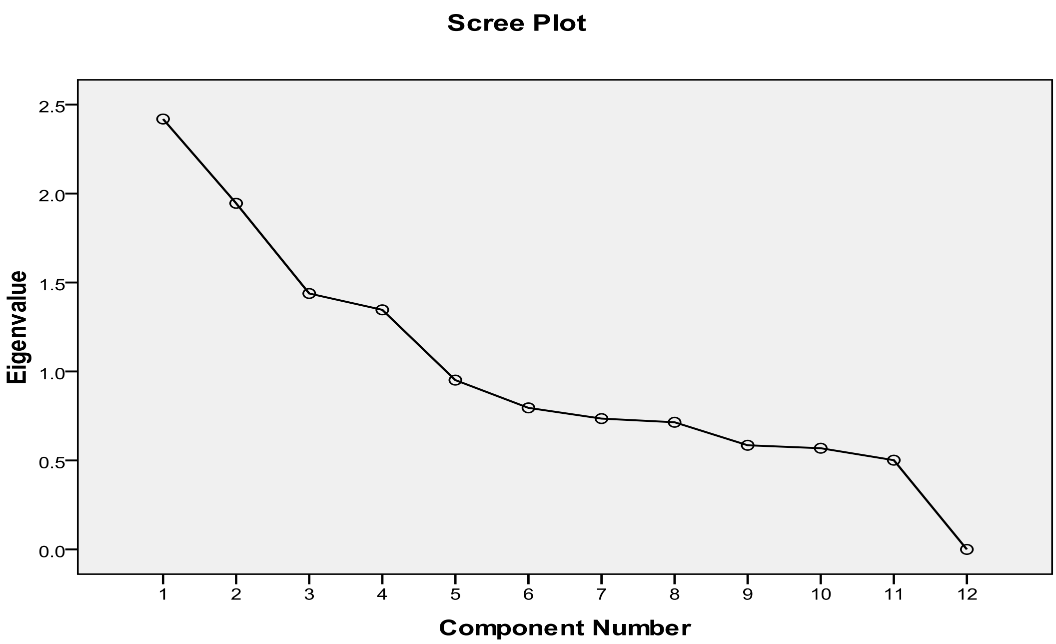

Figure4. *Screw Plot for 12 Parameters in PCA*

Three factors have higher loadings and they are names as Poor Class, Middle Class and Higher Class. . The “Component Matrix” below, gives the factors loadings. This is the central output for factor analysis. The factor loadings are the correlation coefficients between the variables (rows) and factors (columns). The first table 10 gives the unrotated solution and the second the rotated solution. Normally the rotated solution will be appropriate to interpret.

Table11. *Rotated Component Matrix for 12 Parameters*

Rotated Component Matrix			
	Component		
	1	2	3
Has television	**.696**	-.042	-.011
Has refrigerator	**.681**	-.019	-.093
Has electricity	**.590**	-.042	.169
Has motorcycle/scooter	**.584**	-.026	-.012
Has car	**.492**	-.025	-.198
Main floor material	**.454**	-.110	.218
Main roof material	**.335**	.041	.052
Main wall [exterior] material	**.333**	.064	.167
Source of drinking water	-.025	**.998**	-.017
Source of non-drinking water	-.025	**.998**	-.017
Household structure	.076	-.025	**.813**
Number of de facto Household members	.007	-.008	**.810**

Table12. *Component Transformation Matrix*

Component Transformation Matrix			
Component	1	2	3
1	.901	-.391	.187
2	.398	.917	.000
3	-.171	.075	.982

The "Component Transformation Matrix" in table 11 and 12 figure 5 indicates the correlation of the factor before and after rotation. A three dimensional factor-loading plot of the first three factors is shown below. This is the graphical way of presenting the same information as contained in the "Rotated Component Matrix" of factor loading above.

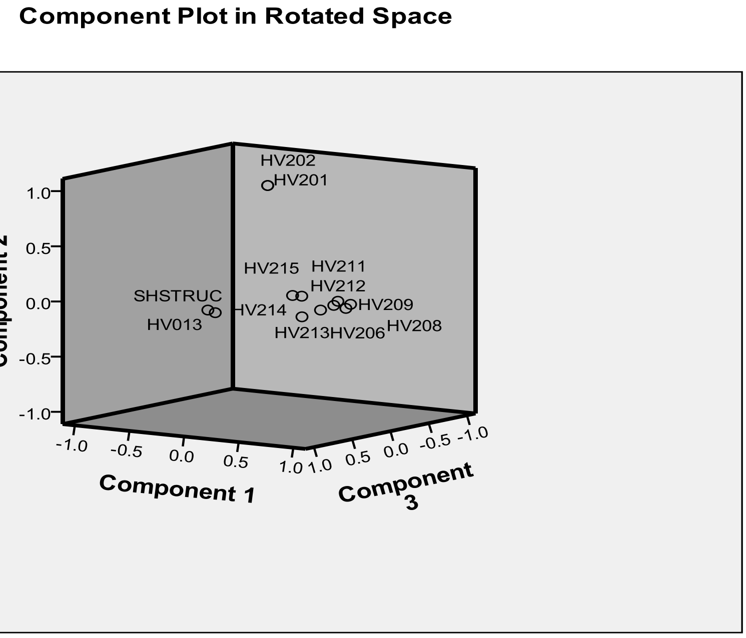

Figure5. *Component Transformation plot in Rotated Space*

4.6. Logistic Regression

In Logistic regression summary table of unweighted cases are in table 13. The Dependent variable results shows that the Non-nuclear family is 0 and Nuclear family is 1..

Table13. *Logistic regression summary statistics of Non-nuclear and nuclear family*

Case Processing Summary			
Unweighted Cases		N	Percent
Selected Cases	Included in Analysis	13806	100.0
	Missing Cases	0	.0
	Total	13806	100.0
Unselected Cases		0	.0
Total		13806	100.0

Dependent Variable Encoding	
Original Value	Internal Value
Non-nuclear	0
Nuclear	1

The various coding schemes for the categorical variables are listed in the Table 14 are shows that the diagonal matrix results are 1 and rest of matrix result is 0.

Table14. *Summary Statistics for Wealth Index and Higher Education*

Categorical Variables Coding						
		Frequency	Parameter coding			
			(1)	(2)	(3)	(4)
Wealth Index	Poorest	142	1.000	.000	.000	.000
	Poorer	556	.000	1.000	.000	.000
	Middle	1653	.000	.000	1.000	.000
	Richer	5232	.000	.000	.000	1.000
	Richest	6223	.000	.000	.000	.000
Highest educational level attained	No education, preschool	2246	1.000	.000	.000	.000
	Primary	3207	.000	1.000	.000	.000
	Secondary	6938	.000	.000	1.000	.000
	Higher	1372	.000	.000	.000	1.000
	Don't know	43	.000	.000	.000	.000

Block 1: Method = Enter

It is observed from the Table 15 shows that all the coefficient are significantly contributing to the model, as the p-value are <0.05

Table15. *Omnibus Test of Model Coefficient*

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	459.905	8	.000
	Block	459.905	8	.000
	Model	459.905	8	.000

From the model summary of Logistic regression in the Table 16, it is observed that the model is fairly good.

Table16. *Model Summary of Logistic Regression*

Model Summary			
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	18385.599	.033	.044

The classification details of the model are shown in Table 17.

Table17. *Classification of Logistic Regression*

Classification Table					
Step 1	Observed		Predicted		
			Household structure		Percentage Correct
			Non-nuclear	Nuclear	
	Household structure	Non-nuclear	7213	695	91.2
		Nuclear	5027	871	14.8
	Overall Percentage				58.6

From the classification Table 17, the researcher observed that 91.2% of the non-nuclear cases are predicted and 14.8% non-nuclear cases are predicted correctly. The overall accuracy of the model to predict the non-nuclear and nuclear of the respondent due to at the household structure is close to 58.6%.

Table18. *The Logistic Regression Model of Household Structure*

Variables in the Equation									
		B	S.E.	Wald	df	Sig.	Exp(B)	95% C. I. for EXP(B)	
								Lower	Upper
Step 1	HV106			308.3	4	.000			
	HV106(1)	-.368	.322	1.30	1	.253	.692	.368	1.30
	HV106(2)	.201	.321	.393	1	.531	1.223	.652	2.29
	HV106(3)	.473	.320	2.18	1	.139	1.604	.857	3.00
	HV106(4)	.740	.324	5.22	1	.022	2.096	1.111	3.95
	HV270			205.1	4	.000			
	HV270(1)	1.238	.180	47.5	1	.000	3.447	2.425	4.90
	HV270(2)	1.042	.093	125.2	1	.000	2.834	2.362	3.40
	HV270(3)	.480	.058	69.27	1	.000	1.616	1.443	1.81
	HV270(4)	.127	.040	10.33	1	.001	1.135	1.051	1.22
	Constant	-.763	.320	5.68	1	.017	.466		

The fitted model is results are in Table 18.

$$\mathbf{Logit(p) = -0.763 - 0.368\ (Praimary) + 0.201\ (Secondary) + .473\ (Higher) + 0.7\ Dont\ Know)}$$
$$\mathbf{+\ 1.238\ (Poorer) + 1.042\ (Middle) + 0.480\ (Richer) +\ 0.127(Richest)}$$

The variable whose corresponding p value are less than 0.05 are poorer,middle,richer and richest which contribute to the response variable household structure of the respondent.

5. Findings and Conclusions

The researcher established from the following results using chi-square test,

1. There is association between household structure and wealth index.
2. There is association between household structure and highest education level attained.
3. There is association between wealth index and highest education level attained

The researcher established from the following results using Principal Component Analysis,

1. Twelve variables are reduced to three factors with 20.15% of the total variation explained by component 1. 36.36% of the total variation explained by component 2. 48.347% of the total variation explained by component 3. Principal Component Analysis (PCA) is extracted score of wealth index are named as Lower Class, Middle Class, and Upper Class.

The researcher established from the following results using Logistic Regression Model,

1. 91.2% of the non-nuclear cases are predicted and 14.8% non-nuclear cases are predicted correctly. The overall accuracy of the model to predict the non-nuclear and nuclear of the respondent due to at the household structure is close to 58.6%.

2. 59.7% of all possible subjects predicted with the response variable nuclear and non nuclear of the respondent due to household structure are classified acceptable discriminations.

Statistically significant association were found between Wealth index, Highest educational level attained and Household structure ($p<0.05$). 12 socio-economic variables are considered to compute wealth index .The first principal component score extracted is the wealth index score which is classified as poor, middle, richer. The relationship between household structure (nuclear /non- nuclear) and wealth index is examined using a Binary logistic regression mode

$$\mathbf{logit}(\boldsymbol{p}) = \boldsymbol{\alpha} + \sum \boldsymbol{\beta}_i \boldsymbol{X}_i$$

Defacto household member and wealth1 (poor) and wealth2 (middle) were significant predictors for household structure with overall classification accuracy of 74.4%. The 95% C.I for odd ratio of defacto household member are [1.817, 1.917], wealth index specified for poor are [0.360, 0.521] and middle are [0.643, 0.820].

REFERENCES

[1] El-Houssainy Abdel Bar Rady, *et. al.* (2011), A Wealth Index of Households Living Conditions in Mauritania/

[2] Diana Lee Ngo (2011), TVs, Toilets, and Thresholds: Measuring Household Wealth Comparably Using an Asset-based Index.

[3] David E. Sahn and David Stifel (2001), Exploring Alternative Measures of Welfare in the Absence of Expenditure Data. Review of Income and Wealth, Series-49, Number-4.

[4] Richard A. and Dean W. (1988). Applied Multivariate Statistical Analysis, Prentice-Hall, Inc. pp. 459-468.

[5] S.S Morris , C Carletto, J Hoddinott, L J Christiaensen (2000), Validity of rapid estimates of household wealth and income for health surveys in rural Africa, Journal of Epidemiology and Community Health, Volume -54, Issue-5, pp.381-387.

[6] Seema Vyas and Lilani Kumaranayake (2006), Constructing Socio-Economic Status Indices: How to Use Principal Components Analysis, Oxford University Press in association with The London School of Hygiene and Tropical Medicine.

[7] World Health Organization (2003). World Health Survey. Geneva, World Health Organization.

[8] Abby Córdova (2008), Methodological Note: Measuring Relative Wealth using Household Asset Indicators, Americas Barometer Insights, No-6. pp. 1-9.

Citation: *Manimannan G, et.al., The Pattern of Wealth Index and Household Structure of Tamilnadu, International Journal of Scientific and Innovative Mathematical Research (IJSIMR), vol. 8, no. 10, pp. 1-13, 2020. Available : DOI: https://doi.org/ 10.20431/2347-3142.0810001*