Binary Logistic Regression Modelling on the Impact of Socio-Economic Variables on Longevity

Ehada Micah Isa & Dr Jonathan A. Ikughur,
Department of Mathematics,
Statistics and Computer Science,
University of Agriculture
Makurdi
wemsmicah@gmail.com

Rilwanu Y. Mohammed
Department of Mathematics/ Statistics,
Nasarawa State Polytechnic,
Lafia, Nasarawa State.

Akyengo Ovye
Directorate of Research and Development,
Nigerian Institute of Leather and Science Technology.

Abstract
Man's desire is to live a long, healthy and qualitative life in the midst of several life challenges that induce stress on him thereby shortening his life. It supposes that when quality of life is attained, there is an artificial creation of an unstressed life that has the tendency of increasing life expectancy. The study assesses the quality of life of civil servant in Nasarawa state using binary logistics regression model in relation to socioeconomic variables namely: literacy, income, family size and location. The Wald criterion demonstrated that gender, family size, locality, literacy and extra income made a significant contribution to longevity (p=0.000). Marital status and income were not significant predictors of longevity, the coefficient of regression β indicate that gender, ' family size, locality, educational qualification, and extra income contributed to the prediction of longevity positively, while both marital status and income contributed negatively to the prediction of longevity and thus are non-significant. Gender equality in employment, increase access to social amenities, increase in service training and spouses who financially support their partner are recommended are measure which could lead to increase life expectancy.

Keywords: *Life expectancy, Binary logistic regression model, longevity, civil servant, literacy.*

1.0 Introduction

Man's desire is to live a long, healthy and qualitative life in the midst of several life challenges that induce stress on him thereby shortening his life. This of course is the concern of the governments, international organizations and world bodies among others, to roll out programmes and hence development initiatives that will enhance quality of life.

Quality of life can be viewed in diverse ways Poon *et al*, (2010) sees it as satisfaction with experiences. This supposes that when quality of life is attained, there is an artificial creation of an unstressed life, that have the tendency of increasing life expectancy otherwise longevity of a person. However, when individuals are dissatisfied with their living, the state of happiness becomes doubtful and certain forms of pressure sets in that reduces one's lifespan. It is

therefore important to investigate those factors that may probably increase or decrease longevity.
The question then is, why do some countries or communities attain longer life while others do not, even as we live in a globalized world. It is therefore of interest to study the quality of life of Nigerian civil servant using Nasarawa State Civil Service as a case study, utilize binary logistics regression model in relation to socio-economic variables namely: literacy, income, family size, gender and location with the view of proffering policy direction for improving the quality of life of the work force, which is the driving force of government.
The study area is Nasarawa State, Central Nigeria with Lafia as its capital .It is the capital city of Nasarawa state and has a population of 330,712 inhabitants (2006 census figure) laying on latitude and longitude of 8.48333◦ and 8.51667◦, with inhabitants who are basically farmers, civil servants and partly business men and women.

2.0 Literature Review

Life longevity on the other hand is living longer or more than the average expected age, which can be determined by many factors such as literacy, income and many other factors (UN, 1991) of course, this study investigates the impact of socio-economic factors on longevity of Nigerian workers, using workers in Nasarawa state civil service as a case study. It utilizes binary logistic regression to predict the quality of life of the people in the study area where, the World Health Organization in 2015showed that male lived for 53 years and females live for 55 years while the average age for a typical Nigerian is 52 years, noting that the life expectancy of Nigeria raised from 46 years in 1990 to 52 years in 2015, and the life expectancy of the world is 71 years. Striving for maximum potential age is the goal of longevity of life which can be realised by creating healthy behaviours and attitudes among others which are key issues advocated by various stakeholders.
Rogers and Wofford (1989) identified the factors affecting longevity developing nations and these includes lack of attention in the social sector, like health, education, sanitation, environmental management and others including changing expenditure structure that affect poverty rate, high illiteracy, easy access to clean water and better nutrition which will positively impact life longevity.

Barro and Sala-I-Martin (1992) worked on the impact of economic growth on life longevity observed that economic growth has great impact on life longevity while Rogot *et al* (1992) investigated the main determinant of life longevity of white female and male in the United States by family income, education, employment status, and they concluded that life longevity varies with mean years of schooling. Gulis (2000) in his studies of the determinant of life expectancy of 156 countries of the world concluded that income per capita, public health, spending, safe drinking water are the main determinant of life longevity.
Heir *et al* (2013) in their work on life longevity among initially healthy middle aged men looked at how various lifestyle factors in midlife predict longevity, they studied longevity in relation to smoking status, body mass index and physical fitness, in 821 healthy men between 51 and 59 years of age, using logistic regression model and concluded that lifestyles variable appears to be strong and independent predictors of life longevity in initially healthy middle men.
This study employ the binary logistic regression model, for analysis and predicting outcome in this study, because it is well suited for describing the relationship between categorical outcome variables (Cabrera, 1994).

3.0 Methodology

The research work is aimed at identifying the socio-economic variables that account for “not living above 52 years” which is obviously the life expectancy of Nigeria (WHO 2015).

Relevant methods relating to logistic regression especially the binary logistic regression models are presented here. The method of data collection is secondary source, where questionnaires were administered to 350 civil servants in Lafia, Nasarawa state whose relations are deceased but have been with the state service. Sample size is determined using the approach by Preduzzi *et al* (1996) which suggested a number of cases to be included and this is given by

$\boldsymbol{n} = \frac{\mathbf{10}\boldsymbol{k}}{\boldsymbol{p}}$. where p is the smallest of the proportion of negative or positive cases in the population, k the number of independent variables. The study assumes the smallest p to be 0.2, while k is equal 7 which will give us a sample size of 350.

The distribution of the samples was done by selecting 30% of the total local government Area (5 local government area) out of the 14 local government areas in the state. Using random numbers digits. The selected areas are shown below and are therefore stratified and sample size defined in proportion to the size of the station.

Table 1: Distribution of persons per Local Government Area

Random Numbers	LGA	Civil Servant	Numbers administered
11	Obi	1523	74
09	Nasarawa Eggon	1716	80
08	Lafia	2730	129
07	Kokona	532	25
10	Nasarawa	888	42
Total		7839	350

Source: Nasarawa State planning Commission 2016

Model Specification.

Consider the model of the form

$$Logit\ \mathrm{y_i} = \frac{\mathbf{1}}{\mathbf{1}+e^{-}(\boldsymbol{\beta_0} + \boldsymbol{\beta_1}\,\boldsymbol{x_1}+\boldsymbol{\beta_1}\boldsymbol{x_2}\ + \ldots+\boldsymbol{\beta_q}\,\boldsymbol{x_q})} \qquad \textbf{(1)}$$

and

$$Y_i = \begin{cases} 0 & \text{if age of respondent is less than 52 years} \\ 1 & \text{if otherwise} \end{cases}$$

$$x_{1i} = \begin{cases} 0 & \text{if respondent is not married} \\ 1 & \text{if otherwise} \end{cases}$$

$$x_{2i} = \begin{cases} 0 & \text{if respondent locality is rural} \\ 1 & \text{if otherwise} \end{cases}$$

$$x_{3i} = \begin{cases} 0 & \text{if respondent is having secondary and below} \\ 1 & \text{if otherwise} \end{cases}$$

$$x_{4i} = \begin{cases} 0 & \text{if respondent income is} \le 40{,}000 \\ 1 & \text{if otherwise} \end{cases}$$

$$x_{5i} = \begin{cases} 0 & \text{if respondent is female} \\ 1 & \text{if male} \end{cases}$$

$$x_{6i} = \begin{cases} 0 & \text{if respondent family size is} >4 \\ 1 & \text{if respondent family size is} \le 4 \end{cases}$$

$$x_{7i} = \begin{cases} 0 & \text{if respondent has no extra income} \\ 1 & \text{if respondent has extra income} \end{cases}$$

Equation 2 is called a logistics response function for any value of $x_{1,} \ldots, x_q$ The commonest measure is the odds of belonging to a class, say, class one(1) and is defined as the relation of the probability of belonging to class 1 to the probability of belonging to class 0. Mathematically

$$\text{Odds} = \frac{\boldsymbol{p}}{\mathbf{1}-\boldsymbol{p}} \qquad \textbf{(2)}$$

While the probability of belonging to a class 1 is

$$\rho = \frac{odds}{1+odds} \qquad (3)$$

The relationship between the odds and the probability is connected by

$$odds = e^{\beta_0 + \beta_1 x_1 + \beta_1 x_2 + \cdots + \beta_q x_q} \qquad (4)$$

Now taking a log to base on both sides, we get the standard formulation of a logistics. Model given as

$$\log(odds) = \beta_0 + \beta_1 x_1 + \beta_1 x_2 + \ldots + \beta_q x_q$$

So that $= \sum_{p=0}^{q} x_{iq}\, \beta_{q}, \quad i=1,2\ldots n$ (5)

$$Y_i = log \frac{p}{1-p}$$

The model assumes the following:

i. It does not require a linear relationship between dependent and independent variable. Hence it should yield a prediction in [0,1] which is not realized in linear regression model.

ii. The dependent variable does not need to be metric and not necessarily a multivariate normal

iii. The variance of y is constant across all classes since y follows a binomial distribution;

iv. Thee independent variable and log odds are linear and;

v. It requires quite a large sample for the purpose of this study,

Further assumption includes:

The cut-off income requirement for a household in Nasarawa state to live a happy and healthy life is a minimum of N40,000 per month, Thus, a sampled household with income <N40,000 per month is assumed to fall into the dichotomous class 0 and income ≥N40,000 per month is assumed to fall into the dichotomous class 1.

A household with a family size of ≤ 4 (in line with National Population policy) will live happier for the given income and belonging to dichotomous class 1 and family size > 4 for a given income is assumed to be confronted with challenges that reduced happiness, hence it belong to dichotomous class 0

Apart from following the binary logistic regression model, other important statistics for making inference in this study include the Overall model evolution, Statistical test for individual predictors, Goodness of fit statistics.

Results and Discussion

The following result from the study are presented table 1 to 5. Table 1 gives the frequency of independent variables. It acts as an important reminder of which categories, were coded as the baseline for each of our independent variables, and it also gives the frequency of each categories as they occur in the analysis. Thus 115 persons were without extra income and 235 have extra income, 226 respondents have a family size greater than 4 while 124 have family size less than or equal to 4. The rural area has 86 respondents while 264 persons were urban dwellers. Married respondents were 264 and 122 respondents were singles, 28 respondents had income less than 40,000.00 naira per month while 322 persons earns 40,000.00 naira and above per month, with 137 having lower qualifications and 213 having higher qualifications; 130 of the respondent are females while 220 are male respondent.

Table 2 in the appendix evaluates the null model, a model with only the constant in the equation. This indicates that the model provides 70.6% accurate classification. However, for the full model (See table 3) where the predictors are included, the binary logistic regression estimates the probability of an event occurring, that is, living more than 52 years. Here the model correctly classified by 85.7% of all the cases which is an improvement over the null base line model. The specificity of the model is the percentage of people with age less than 52, which is 93.9 %.

Table 4: Model summary of the full model

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	253.968[a]	.385	.548

Table 4 provides the –2loglikelihood and pseudo R^2values for the full model. The –2loglikelihood (253.968), is compared to the -2loglikelihood of the null model (424.161) which shows that the new model (with the explanatory variables) significantly provides a better fit than the null model, because there is significant decrease in -2loglikelihood.The R^2values tells us approximately how much variation in the outcome is explained by the model while Nagelkerke's R^2 suggests that the predictor explain roughly 55% of the variation in the longevity.

Table 5: Variables in the equation for the Binary Logistic model

		β	S.E.	Wald	df	P. Values	Exp(β)	95% C.I. for Exp(β)	
								Lower	Upper
Step1[a]	X_1	3.023	.453	44.628	1	.000	20.551	8.446	49.889
	X_2	4.024	.538	55.916	1	.000	55.931	19.480	160.595
	X_3	1.519	.394	14.864	1	.000	4.565	2.110	9.880
	X_4	-.647	.397	2.660	1	.103	.524	.878	1.139
	X_5	1.657	.356	21.687	1	.000	5.242	2.610	10.527
	X_6	-1.052	.659	2.544	1	.111	.349	.096	1.272
	X_7	1.744	.374	21.747	1	.000	5.722	2.749	11.910
	Constant	-6.619	.736	15.736	1	.000	.001		

a. Variable(s) entered on step 1: X_1, X_2, X_3, X_4, X_5, X_6, X_7.

Table 5 gives the contribution of each predictor in the full model and the variables with P-values less than 0.05 are deemed to contribute significantly to the predictive ability of the model (Wright, 1995). Therefore there are 5 variables namely (gender (X_1) with p =.000, family size (X_2) with p =.000, locality(X_3) with p =.000, educational qualification (X_5) with p =.000, and family size (X_7) that are significant having as such the p=.000 all have effect on the probability of living above age greater than 52 years. However income(X_6) and marital status (X_4) with p =0.111 and 0.103 respectively have no significant impact on the probability of living above age 52 years. The results support the claims by Rogot *et al* (1992) whose work observed that literacy has a great impact on longevity. Gulis (2000) actually declared that income is one of the main determinants of life longevity which however contradicts the result of this work which identifies income as no significant impact on life longevity. Rogers and Watford (1989) also discovered that literacy rate and income plays an important role in determining life longevity. In computing the probabilities of cases falling into a specific category, the B values for X_1=3.023, X_2=4.024, X_3=1.519, X_5=1.657 and X_7=1.744 are all significant and positive, indicating that increasing the influence will increased odds of achieving longevity (that is living above average age of 52 years).However, β coefficient for X_4=-0.647 and X_5= -1.052 are not significant and negative, indicating that decreasing their influence will decrease odd of achieving longevity to some extent. The Exp (β) values otherwise known as odd ratio for each of our independent variables shows increase or decrease if the ratio is greater than one or less than one respectively. [Tabachnick and Fidell (2001)]. Therefore Exp (β) values from table 5 has (X_1)=20.551, (X_2)=55.931, (X_3)=4.565, (X_5)=5.442 and (X_7)=5.722 showing that they are clearly greater than one indicating that they have higher odds for the predictors namely Gender (X_1,) family size (X_2), locality (X_3)' literacy (X_5), extra income (X_7) and the outcome variable longevity that is living above age 52. The values for gender (X_1) =20.551 and family size (X_2) have greater odds of predicting the outcome variable age greater than 52 year than other predictors while marital status (X_4)=0.241 and literacy (X_6)=0.096 clearly indicates that, they have a weak relationship with the outcome variable "living above 52 years".

The 95% confidence interval encompasses the true value of the odd ratio, therefore the confidence interval for gender (X_1) with odd ratio (20.551) ranges from [8.466 to 49.889] family size (X_2) with odd ratio (55.931) ranges from [19.480 to 160.595]. The two variables have more stronger relationship than all other variables as such the 95% confident that the actual value of odd ratio for gender (X_1) in the study population lies somewhere between [8.466 and 49.889].

Concluding Remarks

A binary logistic regression was performed to ascertain the impact of some socio-economic variables namely gender, family size, locality, marital status, educational qualification, income and extra income on longevity using Nasarawa state civil servants as a case study. The filled binary regression model was statistically significant with chi-square (χ^2) = 170.193, $p< 0.05$ and df =7. The model explained 55.0% (Nagelkerke R^2) of the variance in longevity (that is living above age 54.5 years) and correctly classified 85.7 % of cases. The Wald criterion demonstrated that gender, family size, locality, literacy and extra income made a significant contribution to longevity (p=0.000). Marital status and income were not significant predictors of longevity. Exp (β) values indicate that gender, family size, locality, educational qualification and extra income have strong relationship with longevity while marital status and income had a very weak relationship with longevity. The coefficient of regression β indicate that gender (X_2)family size (X_2)locality (X_3)educational qualification(X_5)and extra income (X_7)contributed to the prediction of longevity positively, while both marital status (X_4)and income (X_6)contributed negatively to the prediction of longevity and thus are non-significant. Specifically urban dwellers with higher qualifications with income >40,000 have higher probability of living above 52 years than rural dwellers with the same attributes.

6.2 Recommendation

1. Gender equality should be promoted among the civil servant and should be considered in times of employment.
2. There should be increase in awareness on family planning and child spacing to reduce the number of children and dependents to be catered for by a family to reduce the burden of provision.
3. Increase access to social amenities in the rural area will enhance the quality of life in the rural area and in turn increase longevity.
4. There should be increase in service training to help the workforce increase its knowledge and acquired more skills to discharged their responsibility and increase their quality of life.
5. The need to have a spouse whom is financially supported should be encourage and also civil servant should be encourage to invest in activity that will bring extra income.

The outcome here which suggest that income is not important predictor of longevity is at variance as it known interactive. Could this be a realization of the fact that the current minimum wage is not predicting happiness that is required to elongate lives of families in Nasarawa State.

Reference

Adedayo O.A., (2015) Longevity in Nigeria: what and what really matters. *Journal of Econs. and sus dev.: vol 6:no 4. Pp* 142-155.

Agressti, A.1996. An Introduction to Categorical Data Analysis. John Wiley and Sons Inc. Hoboken, New Jersey.

Agresti, A. 2002 categorical data analysis (2nd Edition). John Wiley and son's Inc, Hoboken, New Jersey.

Aromolaran, Adeyemi, D., Oyeyinka, Isaiah K., Olukotun, Oluseyi Okon, Benjamin E., (2003) "Binary Logistic Regression of Student Academics Performances in Tertiary Institutions in Nigeria by Socio Demographic and Economic Factors." *International Journal of Engineering Science and Innovation Technology (IJESIT) Vol. 2, Issue 4.*

Babubhai V.S, Belt G.B (2003) "Hosmer-Lemeshow goodness at fit for survey data" joint statistical meetings – section on survey research methods.

Barrot, R.J and X.I Sala-I- Martin, (1995) "Economic Growth." McGraw-Hill, New York.

Cabrera, A. F. (1994). Logistic regression analysis in higher education: An applied perspective. Higher Education: Handbook of Theory and Research, Vol. 10, 225-256.

Case. K, Fair, R., (2007) "Principles of Economics" Upper Saddle River, NJ Pearson education. p.54

Chao-ying, J.P, Kuk, L.L, Garry M.I (2002) "An Introduction to Logistic Regression Analysis and Reporting". *Journal of Education Research September / October 2002 (vol. 96 (No. 1)*

Cox, D.R., (1958) "The Regression Analysis of Binary Sequences (With Discussion)" J Roy Stat Soc B. 20:215-242.

D. E. Beauvoir, Simone D., (1953) "The Second Sex." Trans.H.M, Parshley. New York.

Dorothy N. Ononokpono (2013) "Maternal Health Care in Nigeria: Do Community Factors Moderate the Effects of Individual Level Education and Ethnic Origin".

Educational Development Center Inc Literacy for Adults. (2015).

Gavrilov, L.A and Gavrilova, N.S 1991. The Biology of Life Span: A Quantitative Approach. Harwood Academic Publisher, New York.

Gujarati, D.N (2004) Basic Econometrics. 4th Ed. New York: Tata Graw–Hill Publishing Co. Ltd.

Gulis G. (2000) "Life Expectancy is an Indicator of Environment Health. *European Journal Epidemiology 2000 Feb; 16(2); 161–165.*

Healey, L.M (2006): Logistic Regression: An Overview, Ester Michigan college of Technology.

Hill, M.A and E.M. King. 1995. "Women's Education and Economic Well-Being", Feminist Economics, (2):21 – 46.

M. Strano, B.M., Colosimo (2006) "Logistic Regression Analysis for Experimental Determination of Forming Limit Diagrams, *International Journal of Machine Tools and Manufacture. 46(6).*

M.B., Skrifvars, V.Pettila, P.H. Rosenberg, M. Castren (2003) "A Multiple Logistic Regression Analysis of in –Hospital Factor Related to Survival at Six Months in Patients Resuscitated from out Of Hospital (Ventricular Fibrillation).

Macfarlane S. M. Racelis and F. Mulimusiime, 2000. "Public Health in Developing Countries". Lancet 356(2)841 – 846.

Menard, S 2002. Applied Logistic Regression Analysis (2nd Edition). Thousand Oaks, Casage publication series quantitative application in the social services No: 106 (1st Edition) 1995.

Natali Mery S.A., Arminda L.S., Clareci S.C., Waleska T.C. (2007). Ordinal logistic regression models: Application in quality of life studies. *Caude publica, Rio De Janeiro,24 sup 4:5581-5591.*

Ogunfiditimi, F, Oguntade, E (2014) "Logistic Regression: a Paradigm for Dichotomous Response Data". *The International Journal of Engineering and Science Issue vol. 3 pages 01- 05 2014.*

Palei,S.K.,Das, S.K.(2009) "Logistic Regression Model for Prediction of Roof Fall Risk in Board and Pillar Working in Coal Mines"; An approach Safety Science.47(88).

Poon L.W, P.Martin , A.J Bishop (2010) "Understanding Centenarian Psychosocial Dynamics and their Contributions to Health and Quality of life" Current Gerontology and Geriatrics Research, vol. 2010, article 10 680657, 13 pages.

Resolution of the all Union Executive Committee and the Soviet People Commissars of September 1924. General Statue on Urban and Rural Settlement.

Rogers, R.E and S. Wofford, 1989. "Life Expectancy in Less Developed Countries," J. Dev. Area, 41(2): 185 – 204.

Rogot, E., Sorlic, P.D, and Johnson, N.J (1992). Life Expectancy by Employment Status Income and Education in the Nation's Longitudinal Mortality study. Public Health Reports, 107, 457 – 461.

Saha G. (2011): Applying Logistic Regression Model to the Examination Results Data in: *Journal of Reliability and Statistical Studies vol. 4 Issue: 105-117*

Scott A.C. (2002). "Maximum Likelihood Estimation of Logistic Regression Models: Theory and Implementation." www.czep.net/contact.html.

Sen, A. 1999. "Economics, Policy and Equity an Overview". In: Tanzi, V. Chu, K. Gupta, Second edition, Economic Policy and Equity. International Monetary Fund, Washington DC, PP: 28 – 42.

United Nation 1991. Handbook of Vital Statistic Systems and Methods, Volume 1, (1991): Legal Organizational and Technical Aspects, United State Studies and Methods, Glossary .series F , No. 35United Nation publication. New York.

Walker, S.H., Duncan, D.B. (1967) "Estimation of the Probability of an Event as a Function of Several Independent Variability." Biometrika 54:167-178.

White, J. (1991). "Dynamics of Family Development in a Theoretical Perspective." New York Guilford Press.

World Health Organization (2006). "Constitution of the World Health Organization." Basic Documents: Forty-Fifth edition.

http://www.apa.org/pi/ses/resources/publications/age.aspx

http://www.worldlifeexpectancy.com/nigeria-life-expectancy

APPENDIX

Table 1: Categorical variables coding of the Binary Logistic model

		Frequency	Parameter coding (1)
Gender	Male	220	1
	Female	130	0
Income	Less than 40,000	28	0
	Greater or equal to 40,000	322	1
Educational Qualification	Secondary School and above	213	1
	Below Secondary school	137	0
Marital Status	Single	122	0
	Married	228	1
Locality	Rural	86	0
	Urban	264	1
Family Size	Greater than 4	226	0
	Less than or equal to 4	124	1
Extra Income	Without extra income	115	0
	With extra income	235	1

Table 2: Classification table for the baseline model

Observed			Predicted		
			Age		Percentage Correct
			Less or equal to age 52	Greater than age 52	
Step 0	Age	Less or equal to age 52	247	0	100.0
		Greater than age 52	103	0	.0
	Overall Percentage				70.6

a. Constant is included in the model.
b. The cut value is .500

Table 3: Classification table for the full model

Observed			Predicted		
			Age		Percentage Correct
			Less or equal to age 52	Greater than age 52	
Step 1	Age	Less or equal to age 52	232	15	93.9
		Greater than age 52	35	68	66.0
	Overall Percentage				85.7

The cut value is .500