Original Research Article

ASSESSMENT OF THE IMPACT OF SOME SOCIO - ECONOMIC VARIABLES ON WORLDWIDE AVERAGE LIFE EXPECTANCY: A LOGIT APPROACH

Abstract

Life Expectancy is a veritable indicator for measuring the total wellbeing of a country. In this paper, the impact of some socio-economic variables on life expectancy. These variables namely: access to electricity, GDP per capita, unemployment total labour force, unemployment of youth total, health expenditure per capita, and adjusted national income common for all countries. Using the data of life expectancy for year 2020 and the Binary Logistic Regression model, result revealed that access to electricity and health expenditure per capita positively impacted on life expectancy. Health expenditure per capita was significant at $\alpha = 0.05$. Countries who have Access to Electricity are 1.113 times (prob=0.53) more likely to attend the worldwide life expectancy. Increase in Health Expenditure per Capita is 1.004 times (prob=0.50) likely to increase life expectancy.

Keywords: Life Expectancy, Logistic Regression, Odd Ratio, Wellbeing, Longevity

UNDER PEER REVIEW

INTRODUCTION

Life expectancy is the average number of years a person is expected to live. It is a key indicator of the health and well-being of a population; reason why the focus of Sustainable Development Goal (SDG) 3 is to "ensure healthy lives and promote well-being for all at all ages". Life expectancy gives the details of health measures of a nation and it is affected by many socio-economic factors (Etikan et al. 2019). According to the Organization for Economic Co-operation and Development (2011), the life expectancy at birth is the predominant approach for measuring the health status of a population as well as evaluating the improvement in health status in each country. Over the past years, efforts by successive governments; especially of the developed countries, international organizations and other stakeholders in providing better working and living environments, increasing education, maternal care and rising per capita income have led to increasing trend in life expectancy globally. Developed countries seem to witness rapid increase in life expectancy while developing countries are witnessing slower growth. For instance, the average worldwide life expectancy for year 2020 was reported as 72.63 years which present a 0.23% increase from 2019 (https://data.worldbank.org). Over the past years, it has been observed that the average life expectancy of many countries fall below the worldwide average. Apart from natural disaster and pandemics which could cause a decline in life expectancy, available information shows that there is a relationship between life expectancy and socio- economic variables. In Ali and Ahmad (2014), it was shown that food production and school enrolment have positive significant impact on life expectancy for Sultanate of Oman. Elsewhere, an investigative analysis using pool data taken on annual level from UN and World Bank databases from 1990–2017 revealed that GDP per capita has positive influence of life expectancy of Macedonia, Serbia, Bosnia and Herzegovina, Montenegro, and Albania and lower values of infant mortality led to higher life expectancy (Miladinov, 2020). Mourad (2021) examined the impact of access to electricity, GDP per capita, mortality rate of children under age five years, physicians and health expenditure per person on life expectancy for One Hundred and Thirty-eight countries using multiple linear regression model. Results obtained revealed that a unit increase in the number of deceased children led to a decrease of about 2.12 months in life expectancy at birth and health expenditure per person had a positive impact on life expectancy at birth. For more studies on the impact of socio-economic variables to life expectancy, interested readers are referred to (Starfield, 2012; Mackenbach et al., 2017; Kabir, 2008; Peter, 2010; Herbst et al., 2012; Mirowsky and Ross, 2000).

The relationship between life expectancy and per capita income is certainly not assured at aggregate level in comparison to developing countries such as Costa Rica and Cuba, which are considered to have high health status. Conversely, the USA is among the industrialized countries besides being the wealthiest country in terms of GDP per capita, but it was ranked to have lower health outcomes compared to many others. Sede and Ohemeng (2015) wrote on socio-economic determinants of life expectancy in Nigeria (1980 – 2011). In their study, it was found that, the conventional socio-economic variables such as per capita income, education and government expenditure on health considered to be highly effective in determining life expectancy of developing countries are not significant in the case of Nigeria.

UNDER PEER REVIEW

The study however suggests that, life expectancy in Nigeria could be improved if attention is given to quality of government health expenditure, unemployment and measures to halt the depreciation of the Nigerian Naira against major foreign currency. Their results suggested that unemployment and inflation are the main economic factors that influence life expectancy negatively. There is still need for continuous investigation in this direction.

In this paper, the impact of some socio-economic variables on 2020 worldwide life expectancy is investigated using the binary logistic model. Socio-economic variables that are common to all categories (developed, developing and under-developed) of countries are considered. The probability of attaining the average life expectancy given these socio-economic variables will also be obtained.

MATERIALS AND METHODS

In this section, we outline the statistical methods based on the theory of Binary Logistic Regression. Interested readers may refer to Rashid (2008) for more information on the theory and application of the Binary Logistic Regression.

Binary Logistic Regression Model

Let Y represent a binary response variable;

$Y_i = 1$; if it is up to world average life expectancy and above.

$Y_i = 0$; if it is below the world average life expectancy.

That is, Y_i take either the value 1 or the value 0 with probability $\pi_i(x), 1 - \pi_i(x)$ respectively, where $\pi_i(x)$ is the mean of the binary variable representing the conditional probability $P(Y_i = 1/\text{x})$ and $x = (x_{io}, x_{i1}, x_{12}, \cdots, x_{ik}) \in \mathbb{R}^{k+1}$ is a vector of $K + 1$ explanatory variables.

Model:

$$\pi_i = \Pr\{Y_i = 1/ X_i = x_i\} \quad (1)$$

$$\pi_i(x) = \frac{\exp(\beta_o + \beta_i X_i)}{1+\exp(\beta_o + \beta_i X_i)} \quad ; \; i = 1,2,\cdots,k \quad (2)$$

where X_i are independent variable.

$\beta_0$and β_i; $i = 1,2,\cdots k$, are unknown parameters.

UNDER PEER REVIEW

The specific transformation of $\pi_i\,(x)$ is called the Logit transformation, it is given by Logit

$$(\pi_i\,(x)) = \, ln\, \frac{\pi_i\,(x)}{1-\pi_i\,(x)} \quad (3)$$

The assumptions of Binary Logistics Regression model are:

i. Logistics regression does not assume a linear relationship between the dependent and independent variables.

ii. The dependent variable must be a dichotomy (two categories).

iii. The independent variable need not be interval, no normally distributed, no linearly related, no equal variance within each group.

Maximum Likelihood Estimation of Binary Logistics Regression Model Parameters

To find the MLE for β, the likelihood function $L(\beta)$ is defined as:

$$L(\beta) = \prod_{i=1}^{N} \pi_i(x)^{y_i}(1-\pi_i(x))^{n_i-y_i} \quad (4)$$

$$= \prod_{i=1}^{N} \left(\frac{e^{\sum_{k=0}^{K} \beta_k x_{ik}}}{1+e^{\sum_{k=0}^{K} \beta_k x_{ik}}}\right)^{y_i} \left(1-\frac{e^{\sum_{k=0}^{K} \beta_k x_{ik}}}{1+e^{\sum_{k=0}^{K} \beta_k x_{ik}}}\right)^{n_i-y_i} \quad (5)$$

$$\prod_{i=1}^{N} \left(e^{y_i \sum_{k=0}^{K} \beta_k x_{ik}}\right)\left(1+e^{\sum_{k=0}^{K} \beta_k x_{ik}}\right) \quad (6)$$

Taking the natural log of (3.11) yields the log likelihood function

$$Log_e L(\beta) = \sum_{i=1}^{N} \left[y_i(\sum_{k=0}^{K} \beta_k x_{ik}) - n_i log_e \left(1+e^{\sum_{k=0}^{K} \beta_k x_{ik}}\right)\right] \quad (7)$$

To find the critical points of $Log_e L(\beta)$

$$\frac{\partial Log_e L(\beta)}{\partial \beta_k} = \sum_{i=1}^{N} [y_i x_{ik} - n_i \pi_i(x) x_{ik}] = 0 \quad (8)$$

The results obtained in Equation (8) if it exist gives the MLE of β_i; $\quad i = 0,1,2,\cdots k$

Chi-Square Goodness of Fit Test

To test for the goodness of fit of the model, the $-2\; Log - likelihood$ Null and Full models are usually employed.

The Null model $-2\; Log - likelihood$ is given as:

UNDER PEER REVIEW

$-2 * ln(L_0)$ (9)

where L_0 is the likelihood of obtaining the observations if the independent variables have no effect on the outcome.

The Full model $-2\ Log - likelihood$ is given as;

$-2 * ln(L)$ (10)

Where L is the likelihood of obtaining the observations if the independent variables have effect on the outcome..

The difference from Equations (10) and (9) yields a chi-square statistic. The Chi-Square statistic is tested at 0.05 level of significance.

The Hosmer-Lemeshow (H-L) Test

The Chi-square statistic for the H-L test is given as:

$$\chi^2_{HL} = \sum_{g=1}^{G} \frac{(O_g - E_g)}{E_g(1 - E_g/n_g)} \quad (11)$$

where $O_g, E_g\ and\ n_g$ are the observed events, expected events and number of observations for the g^{th} decile group and G the number of groups with degree of freedom as $G-2$. A large value of chi-square with small p-value < 0.05 indicates poor fit while a small chi-square value with p –value closer to 1 implies a good model fit.

Wald Statistics

The Wald's statistic tests the significance of model parameters. Thus it determines whether or not an independent variable stays in the model as it tests if the associated model parameter differs significantly from zero. The Wald's statistic is computed as the regression coefficient divided by its standard error squared:

$$\left(\frac{\beta}{SE}\right)^2 \quad (12)$$

Where $\beta = the\ associated\ regression\ parameter, SE = standard\ error.$

If p-value is less than the $\alpha = 0.05$, then we have evidence to conclude that the independent variable differ significantly from zero and should be included in the model.

Odds Ratio (O.R.)

Odd Ratios Re-writing equation (1) by taking the exponential of both sides of the equation, we have;

$$Odds = \frac{p}{1-p} = e^{\beta_0} e^{\beta_1 X_1} e^{\beta_2 x_2} \ldots e^{\beta_k X_k} \quad (13)$$

UNDER PEER REVIEW

If $O.R > 1$ it implies that the event of interest is more likely to occur as the predictor increases. But if $O.R < 1$, it is an indication that the event of interest is likely to occur as the predictor increases.

The Data

The data used in this work was obtained from World Bank website (www.data.worldbank.org.idicators), which consist of several socio-economic variables and life expectancy data of 70 countries in 2020. Out of these socio-economic variables, only six (6) were used in this paper due to the incomplete information of other variables. The socio-economic variables used are health expenditure per capita, adjusted net national income, access to electricity, unemployment of total labour force, unemployment youth total and Gross Domestic Product per capita (GDP). Using the Forward Stepwise (Likelihood Ratio) procedure the data was analyzed with the help of IBM SPSS version 2020.

RESULTS AND DISCUSSION

In this section, results obtained from the Logistic Regression Analysis of the data are provided. From Table 1, out of 70 countries, 29 (41.4%) countries have life expectancy that is lower that the worldwide average life expectancy.

Figure 1 presents a graph that compares the life expectancy (LE) of individual countries with the worldwide average Life expectancy of 72.63 years. From Figure 1, it is obvious that LE of many countries fall below the worldwide average LE. For instance, countries like Nigeria and Cote divore have LE as low as 55 and 58 years respectively.

Table 1: Frequency Distribution

Classification	Frequency (%)
< 72.63 years	29 (41.4)
≥ 72.63 years	41 (58.6)

UNDER PEER REVIEW

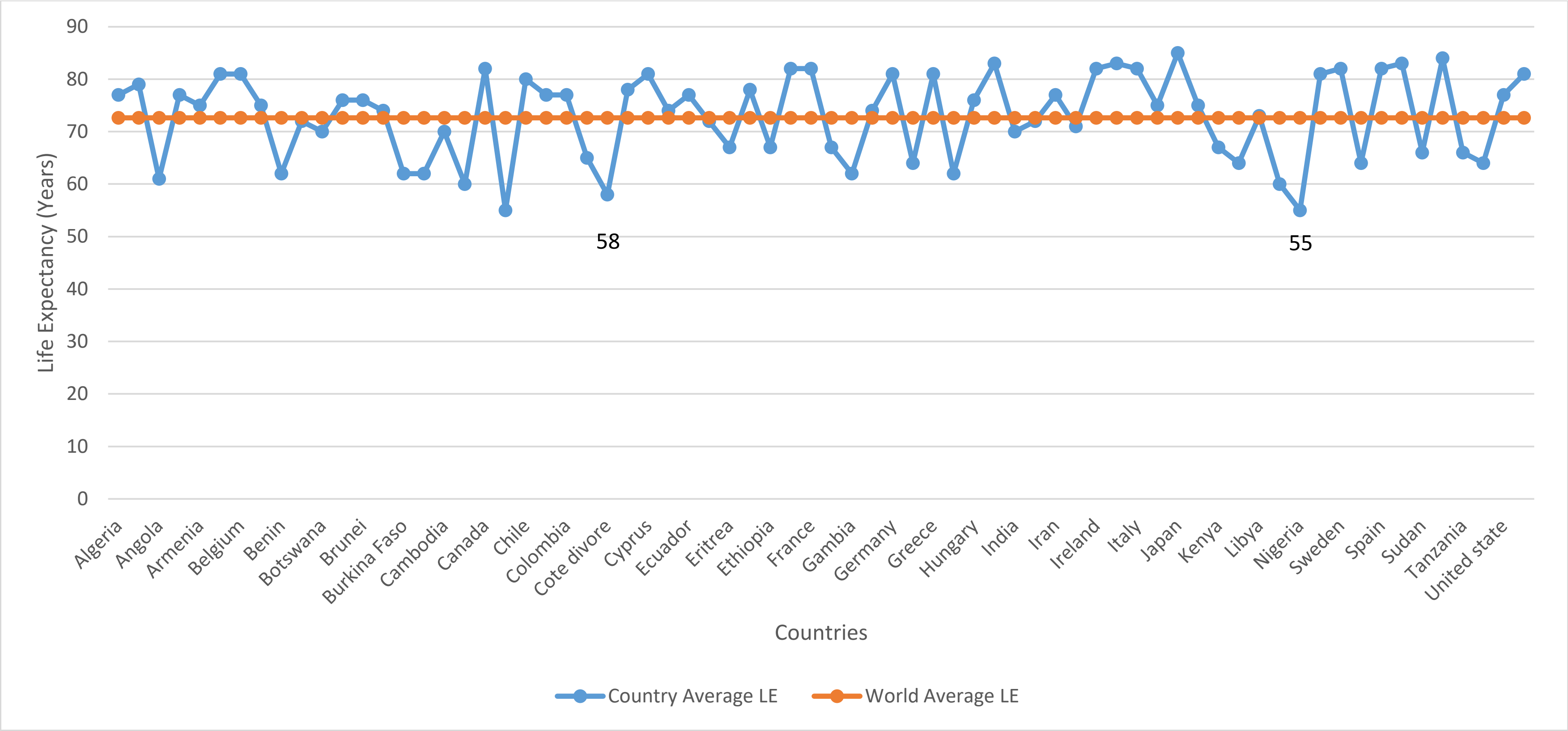

Figure 1: Country Life Expectancy (LE) Versus Worldwide Average Life Expectancy

The following variables were considered in the Logistic regression model (see Equation 2)
Response variable (Y) = Life expectancy

The explanatory variables are:

X_1 = Access to Electricity (%)

X_2 = GDP per Capita

X_3 = Unemployment total of labour force (%)

X_4 = Unemployment of youth total (%)

X_5 = Health expenditure per capita

X_6 = Adjusted National Income

Table 2: Omnibus Tests of Model Coefficients

		Chi-square	Degree of Freedom	Sig.
Step 1	Step	53.467	1	.000
	Block	53.467	1	.000
	Model	53.467	1	.000
Step 2	Step	18.750	1	.000
	Block	72.217	2	.000
	Model	72.217	2	.000

Table 2 presents the result of Omnibus test of model coefficients. The Omnibus test is used to check the adequacy of the model at each step of the analysis. To how the selected explanatory variable at each step improve on the model. If the Chi square value of a later model is significantly less than that of the baseline model, then one can say that the later model explains more variance in the outcome compared the former. In this case, comparing the Chi square value in step 1 and step 2 shows that the Chi square value (18.750) at step 2 is significantly (Sig. Value = 0.00) smaller than the one at step 1 (53.467). This implies that the model in step 2 explains more variance in the outcome compared to model in step 1. The block and model are equally significant with Significant values = 0.00. Note that the optimal results were obtained at step 2. Table 3 further present a summary check on the Logistic regression model. From Table 3, the -2 Log Likelihood value)22.756) in step 2 is less than that of step 1 (41.506) which is indicates an improvement in the Logistic regression model in step 2. The Cox & Snell R Square and Nagelkerke R Square are both summary statistics that shows the amount of variation in the outcome that is explained by the model. In step 2, the Cox & Snell R Square and Nagelkerke R Square indicates that the model explained 64.4% and 86.7% of the variation in the outcome.

UNDER PEER REVIEW

Precisely, Nagelkerke R Square shows that the explanatory variables that were considered in the logistic regression model has explained 86.7% of the variation in life expectancy.

Table 3: Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	41.506[a]	.534	.719
2	22.756[b]	.644	.867

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.
b. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

Table 4 presents another model adequacy check based on the Hosmer-Lemeshow statistic. The significant values in Step 1 and 2 are greater than 0.05 which is an indication that the model makes a good fit. It also shows that there is an improvement in the model at step 2 compared to Step 1.

Table 4: Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	7.254	4	.123
2	7.114	8	.524

Table 5: Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	Access to Electricity	.178	.060	8.850	1	.003	1.195	1.063	1.345
	Constant	-15.894	5.830	7.433	1	.006	.000		
Step 2[b]	Access to Electricity	.107	.060	3.190	1	.074	1.113	.990	1.253
	Health_Exp._Per _Capita	.004	.002	5.611	1	.018	1.004	1.001	1.008
	Constant	-12.965	6.112	4.500	1	.034	.000		

a. Variable(s) entered on step 1: Access to Electricity
b. Variable(s) entered on step 2: Health_Expenditure_Per_Capita.

Table 5 presents results of the logistic regression analysis based on the model at each step. Recall that in the forward stepwise method, the procedure commences with a model that does not include any explanatory variable. At each step, the explanatory variable with the highest score statistic and a significant value less than 0.05 is selected for the model. In this case, the first

UNDER PEER REVIEW

explanatory variable that was added to the model in step 1 was Access to Electricity. Although Access to electricity was significant with value = 0.003, the model still needed some improvement as was explained from Table 3. In step 2, another explanatory variable; Health Expenditure Per Capita that was added to the model is significant with value = 0.018. Based on our model, (Step 2), the variables that greatly explained the variation in counties life expectancy are Access to Electricity and Health Expenditure Per Capita. Access to Electricity and Health Expenditure Per Capita have positive coefficients 0.107 and 0.004 respectively. This means that although Access to Electricity is not significant at step 2, yet it contributes positively to life expectancy. The $Exp(\beta)$ column in Table 5 present the odd ratio (O.R.) which says much about the odds of life expectancy increase or decrease. An O.R. of 1.113 means that those who have Access to Electricity are 1.113 times ($prob = 0.53$) more likely to attend the worldwide life expectancy. An O.R. of 1.004 for Health Expenditure per Capita implies that increase in Health Expenditure per Capita is 1.004 times ($prob = 0.50$) likely to increase life expectancy.

CONCLUSION

Life expectancy is a veritable indicator for measuring the total wellbeing of a nation. This paper examined the impact of six (6) socio-economic variables namely: access to electricity, GDP per capita, unemployment total labour force, unemployment of youth total, health expenditure per capita, and adjusted national income on average life expectancy. Results obtained indicated that access to electricity and health expenditure per capita contributed positively to life expectancy of 2020. Health expenditure per capita was significant at 0.05 level of significance. This means that appropriate health financing has significant positive impact on life expectancy irrespective of country. This result is in support of other research findings which showed that there is a strong positive relationship between access to electricity and economic wellbeing (Stern et al., 2019), which in turn has the tendency to improving life expectancy.

UNDER PEER REVIEW

REFERENCES

Ali, A. and Ahmad, K. (2014) The Impact of Socio-Economic Factors on Life Expectancy for

Sultanate of Oman: An Empirical Analysis. *MPRA Paper* No. 70871, at https://mpra.ub.uni-muenchen.de/70871/

Etikan, I., Babatope, O., Yuvalı, M., and Ilgi, S. (2019). Effects of Socioeconomic Factors on Nigeria's Life Expectancy at Birth. Annals of Biostatistics & Biometric Applications. 1(5): 2019. ABBA.MS.ID.000523. DOI: 10.33552/ABBA.2019.01.000523.

Stern, D. I., Burke, P. J. and Bruns, S. B. (2019): The Impact of Electricity on Economic Development: A Macroeconomic Perspective. Energy and Economic Growth: Applied Research Programme. 1.1: 1-44.

Miladinov, G. (2020). Socioeconomic Development and Life Expectancy Relationship: Evidence from the EU Accession Candidate Countries. *Springer Open* 76 (2). https://doi.org/10.1186/s41118-019-0071-04

Herbst, S., Fayzieva, D. and Kistemann, T. (2008) Risk Factor Analysis of Diarrhoeal Diseases in the Aral Sea area (Khorezm, Uzbekistan), *International Journal of Environmental Health Research,* 18:5, 305-321, DOI: 10.1080/09603120701834507

Kabir, M. (2008). Determinants of Life Expectancy in Developing Countries. *The Journal of Developing Areas*, 41(2), 185-204. http://dx.doi.org/10.1353/jda.2008.0013

Mackenbach J. P., Bopp, D. P. (2017). Determinant of the Magnitude of Socioeconomic'. In Equalities in Mortality: A Study of 17 European Countries. *Health place* 201; 47:44-53.

MOURAD, M. (2021). Impact of Socio-Economic Variables on Life Expectancy: An Empirical Study for 138 Countries. *Journal of Public Administration and Governance, Macrothink Institute,* vol. 11(1), pages 330346-3303.

Mirowsky, J. and Ross, C. E. (2000). Socioeconomic Status and Subjective Life Expectancy. *Social Psychology Quarterly*. Vol. 63, no. 2, 2000, pp. 133–51. JSTOR, https://doi.org/10.2307/2695888. Accessed 8 Sep. 2022

OECD Journal on Development: Development Co-operation Report 2011

Rashid, M. (2008). *Inference on Logistic Regression Models*. Doctoral dissertation, Bowling Green State University.Ohio. http://rave.ohiolink.edu/etdc/view?acc_num=bgsu1214165101

Sede, P.I., Ohemeng, W.(2015). Socio-economic Determinants of Life Expectancy in Nigeria (1980 – 2011). *Health Econ Rev* 5, 2. https://doi.org/10.1186/s13561-014-0037-z

Starfield, B. (2012). Is US health really the best in the world? *The Journal of American Medical Association*, 284(4), 483-485. http://dx.doi.org/10.1001/jama.284.4.483.

UNDER PEER REVIEW