A COMPARATIVE STUDY OF THE MULTIPLE LOGISTIC REGRESSION, LINEAR DISCRIMINANT ANALYSIS AND QUADRATIC DISCRIMINANT FOR ESTIMATING THE MISCLASSIFICATION ERROR RATE OF INFANT BIRTH OUTCOME

Akangoziri O. [1]✉, Okoli C. N. [2]✉

[1,2] Department of Statistics, Chukwuemeka Odumegwu Ojukwu University, Anambra State, Nigeria

DOI: https://doi.org/10.29121/granthaalayah.v8.i9.2020.104

ABSTRACT

This study examined comparison of the Multiple logistic regression, Linear discriminant analysis and Quadratic discriminant in estimating the infant birth outcome and misclassification error rate of birth outcomes with factors of infant mortality in Anambra State, Nigeria. The birth outcomes of interest were the Neonatal death, Still birth and Alive. Secondary source of data were obtained from the records department of General Hospital Onitsha from 2007-2016. The data comprises of Status of infant birth, Mothers parity, Age of mother, Weight of baby, Mothers Education Status, Number of Bookings before gestation and Gestation Age. The data analysis is performed using R-software. The result of the findings from the multiple logistic regression showed that Mothers Education Status (MES) and Booking contributed significantly on the logistic model while factors of Parity, Sex, Age of Mother (AOM), Year, GA and Birth Weight (BW) were found to be insignificant on birth outcomes. Also observed that the misclassification error rate for birth outcome for the said approach is found to be 0.5992 (59.92%). More so, findings of the study equally showed that the prior probabilities of the groups for the linear and quadratic discriminant analysis were 0.228503, 0.40168 and 0.36981 for Alive, Neonatal death and Still birth respectively. Further findings revealed that the Mothers Education Status and Bookings Status have the greatest impact for first and second linear function respectively. In addition, the result of the misclassification error rate for birth outcome using the linear discriminant analysis is 0.5931 (59.31%). The misclassification error rate for birth outcome based on quadratic discriminant analysis is 0.5956 (59.56%). Based on the findings of this study, linear discriminant approach is the best alternative in estimating misclassification error rate of infant birth outcome followed by quadratic discriminant analysis and the least is multiple logistic regression. The findings clearly confirmed that the linear discriminant analysis is the best with misclassification error rate of 59.31%.

1. INTRODUCTION

Infant mortality is a fundamental measure of a country's level of socio-economic development and the quality of life, especially of mothers. Infant mortality has remained a national and global concern and its importance in the socio-economic assessment of the country's development cannot be overstated. Sub-Saharan Africa and South Asia face the greatest challenges in child survival and currently account for more than 80% of deaths of children under the age of five worldwide. Several factors have been identified as contributing to the increase in the infant mortality rate in most developing countries. Studies have shown that there is a close relationship between educational attainment and lower mortality rates.

Although there are vagaries of statistics and estimates of infant mortality for different countries and the world from different sources, the patterns and trends are specifically similar. Among general trends, the global under-five mortality rate fell by almost 47% between 1990 and 2012 (measuring 90 deaths per 1,000 live births in 1990 and 48 in 2012) while the trend in sub-Saharan Africa is likely to increase. Globally, several causes of death in children under five have been noted, including: pneumonia, which accounts for up to 17% of all deaths, complications of premature births which cause about 15 % of child deaths, childbirth complications (10%), diarrhea (9%) and up to 7% due to malaria (United Nations Interagency Group for Estimating Infant Mortality, 2013). In addition, a survey in Bangladesh shows that the infant mortality rate was highest (1.64%) for children of illiterate mothers and lowest (0.54%) for children with educational attainment. of the mother was secondary and higher (Uddin et al., 2009). Educated mothers are more likely than illiterate mothers to provide a healthy environment, nutritious food, and to have a better understanding of reproductive health when designing and providing health care for their children. Literate mothers has been expected to give birth to healthier babies because they themselves tend to be healthier and are likely to experience lower mortality in their children at all ages.

Researchers have continually raised alarm on the high rate infant mortality which still remains a major issue of concern in Nigeria. Also, it is sad to note that the quality of care given during and after birth of an infant is rather rudimentary as there is generally insufficient Government investment in public healthcare. The worst-case scenario is that majority of births do occur in unorthodox facilities and babies only get to hospital after irreparable damage may have occurred. There is need for analytical studies both at the sub-national and household levels in order to produce more reliable empirical evidence to inform the design of policies and programmes that could improve child health outcomes in various countries. Also, some literature noted that it would be methodologically wrong to fit a single-level standard regression model in the analysis of child survival, because of the hierarchical nature of mortality data. Meanwhile, such studies that consider hierarchical structure of mortality data to establish the contextual factors influencing infant and child mortality are limited in Nigeria. However, improving the counting of stillbirths and neonatal deaths is important to tracking Sustainable Development Goal 3.2 and improving vital statistics in low-income and middle-income countries (LMICs). However, the validity of self-reported stillbirths and neonatal deaths in surveys is often threatened by misclassification errors between the two birth outcomes. Study such as Liu et al. (2016) recommend examining the extent of stillbirths being misclassified as neonatal deaths for larger sample size in Malawi or other developing countries. There is need to determine the misclassification error rate associated with reporting stillbirths and neonatal death in Anambra State, Nigeria, hence, the essence of this study.

2. REVIEW OF RELATED LITERATURE

Richardson et al. (2014) stated the reduction in child mortality is necessary in order to attain sustainable development goals. They identified the existence of a major challenge in the procurement of healthcare services by individuals which is determined to a large extent by their level of income. In their study, infant mortality rate, under-five mortality rate and neonatal mortality rate were modeled against household income and controlled for access to anti-natal care, access to safe water and sanitation, neonatal mortality rate, maternal education and household size in Nigeria. The findings of their study revealed that household income has significant effect on neonatal mortality rate in Nigeria but household income has insignificant effect on infant and under-five mortality rates in Nigeria. Also, it was found that household size has significant effect on infant mortality rate and neonatal mortality rate in Nigeria. In addition, findings revealed that access to anti-natal care has significant effect on under-five mortality rate in Nigeria.

Amzat and Adeosun (2014) examined the nature of relationship between infant mortality and some socioeconomic and demographic variables. Also, assessed the proximate covariate that influences the survival of an infant using the 2003 Nigeria Demographic and Health Survey Data (NDHS). They used sequential probity model to examine the relationship between the dependent variables (infant's death and age at death) and predictor variables for both correlated and uncorrelated error terms. The findings of their study showed that in both of the situation with correlated and uncorrelated error terms, infant's being alive or death is positively affected by education, birth order number, duration of breast feeding and negatively affected by both total children born and place of delivery. There exist significant differences among the predictor variables on the probability of infant's death at neonatal and post neonatal period. Also, the correlation between the error terms was found to be significant.

Adetoro and Amoo (2014) stated that despite the global decline in infant mortality rate from 90 deaths per 1,000 live births in 1990 to 48 in 2012, Nigeria is yet to record any substantial improvement. Infant mortality in Nigeria increased from 138 per 1,000 live births in 2007 to 158 per 1,000 live births in 2011 against the Millennium development Goal target of 71 per 1,000 live births. They used data from the Nigeria Demographic and Health Survey (NDHS) 2008 to investigate the predictors of child (aged 0-4 years) mortality in Nigeria. They statistical tool employed were cross-tabulation and binary logistic regression techniques. The findings of their study showed that mortality rate was highest (49.14%) for children of illiterate mothers and lowest (13.29%) among mothers with higher education. Also, the result of the logistic regression analysis revealed that, education of both parents and occupation of mothers were found statistically significant to reduction in child mortality rate. It was equally found that mothers' wealth index, age at first birth and usual of place of residence have substantial impact on child mortality in Nigeria. They concluded that increase in women education could increase age at first birth and mitigate the risk of poor child health outcomes.

Adepoju (2015) examined the differentials in child mortality rate across socioeconomic, demographic and selected health characteristics in rural Nigeria, employing the 2008 National Demographic and Health Survey data. The findings of his study on health attributes and morbidity pattern of mother and child revealed that most of the respondents did not have access to good health facilities and antenatal care. As a result, more than three-quarters of the respondents delivered their babies at home and had less than 24 months birth interval between pregnancies. Results showed that child mortality rate was highest among illiterate mothers, mothers without a source of income, under aged women (less than 20 years) and among fathers whose primary livelihood lie in agriculture. Regional analysis showed that the North-Western zone had the highest child mortality rate followed by the North-Eastern zone, while the South-South zone had the lowest. With respect to health attributes, children delivered at home, who were never breastfed and of multiple births had high mortality rates. Gender differentials showed that the rate of mortality was higher for male than for female children but lowest for children who had been fully immunized and whose mothers were aged between 21 and 30 years.

Jacdonmi et al. (2016) reviewed the trends and patterns of breastfeeding, causes of infant mortality and breastfeeding of infants from birth to six months, followed by appropriate and adequate complementary feeding for two years and above, as a strategic intervention against infant mortality and the need to create awareness about the benefits of breastfeeding. The outcome of their review showed that breastfeeding protects infants from several infections such as diarrhoea, pneumonia, gastrointestinal infections, urinary tract infections, sudden infant death syndrome and others which are probable causes of infant deaths. They noted that as breastfeeding provides adequate nutrition to infants, protects them from diseases and infections, it is a cost-effective method/intervention to reduce infant mortality.

Liu et al. (2016) assessed the extent and correlates of stillbirths being misclassified as neonatal deaths by comparing two recent and linked population surveys conducted in Malawi, one being a full birth history (FBH) survey, and the other a follow-up verbal/social autopsy (VASA) survey. The result of their study found that one-fifth of 365 neonatal deaths identified in the FBH survey were classified as stillbirths in the VASA survey. Neonatal deaths with signs of movements in the last few days before delivery reported were less likely to be misclassified stillbirths (OR = 0.08, p<0.05). It was found that having signs of birth injury has impact on higher odds of misclassification (OR = 6.17, p<0.05).

3. MATERIAL AND METHODS

3.1. METHOD OF DATA COLLECTION

Secondary source of data used in this study obtained from the records department of General Hospital Onitsha from 2007-2016. The data comprises of Status of infant birth, Mothers parity, Age of mother, Weight of baby, Mothers education status, Number of Bookings before gestation and Gestation Age.

3.2. METHOD OF DATA ANALYSIS

The statistical tools used in this study include Multiple logistic regression, linear discriminant analysis and Quadratic discriminant analysis.

3.2.1. MULTIPLE LOGISTIC REGRESSION ANALYSIS

The Multiple logistic regression analysis is a statistical tool used in predicting categorical placement in a dependent variable or the probability of category membership on a dependent variable based on multiple independent variables. The independent variables can be either dichotomous (i.e., binary) or continuous (i.e., interval or ratio in scale). Multiple Logistic regression allows for two or more categories of the dependent or outcome variable (categories such as 0, 1, 2 or -1, 0, 1). Logistic regression uses maximum likelihood estimation to evaluate the probability of categorical membership. Logistic regression does necessitate careful consideration of the sample size and examination for outlying cases. Like other data analysis procedures, initial data analysis should be thorough and include careful univariate, bivariate, and multivariate assessment. Specifically, multicollinearity should be evaluated with simple correlations among the independent variables. Also, multivariate diagnostics (i.e. standard multiple regression) can be used to assess for multivariate outliers and for the exclusion of outliers or influential cases. Sample size guidelines for multinomial logistic regression indicate a minimum of 10 cases per independent variable (Schwab, 2002). Logistic regression is often considered an attractive analysis because it does not assume normality, linearity, or homoscedasticity. Logistic regression does have assumptions, such as the assumption of independence among the dependent variable choices. This assumption states that the choice of or membership in one category is not related to the choice or membership of another category (i.e., the dependent variable). The assumption of independence can be tested with the Hausman-McFadden test (Hoffmann, 2003). Furthermore, multinomial logistic regression also assumes non-perfect separation. If the groups of the outcome variable are perfectly separated by the predictor(s), then unrealistic coefficients will be estimated and effect sizes will be greatly exaggerated.

Suppose we have n independent observation with p explanatory variables. The qualitative response variable has k categories. To construct the logits in the multinomial case one of the categories is considered the base level and all the logits are constructed relative to it. Any category can be taken as the base level, we shall take category k as the base level in our description of the method. Since there is no ordering, it is apparent that any category may be labeled k. Let π_j denote the multinomial probability of an observation falling in the jth category. To obtain the relationship between this probability and the p explanatory variables, $X_1, X_2, ...,X_p$. The general logistic regression model is expressed as

$$\log\left(\frac{\pi_j(x_i)}{\pi_k(x_i)}\right) = \beta_{0j} + \beta_{1j}x_{1i} + \beta_{2j}x_{2i} + \cdots + \beta_{pj}x_{pi} \tag{1}$$

Letting all the $\pi's$ add to unity, (1) will then have the form

$$\log\left(\pi_j(x_i)\right) = \frac{\exp(\beta_{0j} + \beta_{1j}x_{1i} + \beta_{2j}x_{2i} + \cdots + \beta_{pj}x_{pi})}{1 + \sum_{j=1}^{k-1}\exp(\beta_{0j} + \beta_{1j}x_{1i} + \beta_{2j}x_{2i} + \cdots + \beta_{pj}x_{pi})} \tag{2}$$

for j = 1,2, . . . , (k - 1). The model parameters are estimated by the method of maximum likelihood.

3.2.2. LINEAR DISCRIMINANT ANALYSIS

Discriminant analysis is used in situations where the clusters are known a priori. The aim of discriminant analysis is to classify an observation, or several observations, into these known groups. Discriminant analysis is a method used to distinguish between groups of populations Π_j and to determine how to allocate new observations into groups.

Fishers (1936) proposed a linear discriminant function, his discriminant rule was based on a projection $a^\tau x$ such that a good separation can be achieved.

Suppose $y = xa$

is a linear combination of observations, then the total sum of squares of y,

$\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$ can be represented as

$$y^\tau H y = a^\tau x^\tau H x a = a^\tau T a \tag{3}$$

where the centering matrix $H = I - n^{-1} 1_n 1_n^\tau$ and $T = x^\tau H x$

Considering x_j, j=1, ..., J, samples from j populations, the linear combination $a^\tau x$ which maximizes the ration of the between-group-sum of squares to the within-group-sum of squares. The within-group-sum of squares is given by

$$\sum_{j=1}^{J} y_j^\tau H_j y_j = \sum_{j=1}^{J} a^\tau x_j^\tau H_j x_j a - a^\tau W a \tag{4}$$

Where y_i denotes the jth sub-matrix of y corresponding to observations of group j and H_j denotes the (n_j X n_j) centering matrix (Wolfgang and Léopold, 2007). The within within-group-sum of squares measures the sum of variations within each group.

The between-group-sum of squares is given by

$$\sum_{j=1}^{J} n_j\left(\bar{y}_j - \bar{y}\right)^2 = \sum_{j=1}^{J} n_j\left\{a^\tau(\bar{x}_j - \bar{x})\right\}^2 = a^\tau B a \tag{5}$$

Where $\bar{y}_j$ and $\bar{x}_j$ are the means of y_j and x_j and $\bar{y}$ and $\bar{x}$ are the sample means of Y and X. The between-group-sum of squares measured the variation of the means across groups.

The total sum of squares represented in (3) is the sum of the within-group-sum of squares and the between-group-sum of squares.

$$a^\tau T a = a^\tau W a + a^\tau B a \tag{6}$$

To select the projection vector a that maximizes the ratio

$$= \frac{a^\tau \, Ba}{a^\tau \, Wa} \tag{7}$$

The vector a that maximizes (7) is the eigenvector of $W^{-1}B$ that corresponds to the largest eigen value. Hence, the corresponding discriminant rule is

$$x \to \pi_1 \quad \text{if} \quad a^\tau \left\{ x - \frac{1}{2}(\overline{x}_1 + \overline{x}_2) \right\} > 0$$

$$x \to \pi_2 \quad \text{if} \quad a^\tau \left\{ x - \frac{1}{2}(\overline{x}_1 + \overline{x}_2) \right\} \leq 0$$

3.2.3. QUADRATIC DISCRIMINANT ANALYSIS

Quadratic Discriminant Analysis (QDA) is a variant of LDA in which an individual covariance matrix is estimated for every class of observations. QDA is particularly useful if there is prior knowledge that individual classes exhibit distinct covariances. One weakness of QDA is that it cannot be used as a dimensionality reduction technique.

In QDA, $\sum_k$ is required for each class of $k \in \{1, \cdots, K\}$ rather than assuming $\sum_k = \sum$ as it is done in LDA. The quadratic discriminant function is given as

$$\delta_k = -\frac{1}{2}\log\left|\sum_k\right| - \frac{1}{2}(x - \mu_x)^T \sum_k^{-1} (x - \mu_x) + \log \pi_k \tag{8}$$

Since QDA estimates a covariance matrix for each class, it has a greater number of effective parameters than LDA. The quadratic discriminant analysis is quadratic in nature and contains a second order terms.

The classification rule for the quadratic discriminant function:

$$\hat{G}(x) = \arg\max_k \, \delta_k(x) \tag{9}$$

The classification rule is equally similar to LDA since all that is expected is to find the class k which maximizes the quadratic discriminant function.

4. RESULT OF DATA ANALYSIS

4.1. RESULT OF MULTIPLE LOGISTIC REGRESSION ANALYSIS

Table 1: Descriptive Distribution of the Birth Outcome

	Frequency	Percent	Cumulative Percent
Alive	380	22.8	22.8
Neonatal Death	668	40.1	63.0
Stillbirth	616	37.0	100.0
Total	1664	100.0	

Result obtained in table 1 showed that 40.1 % of the birth outcome was neonatal deaths, 37.0% were still births and 22.8% were Alive.

Table 2: Parameter Estimates and model summary

Model	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	1.149	0.1529	7.516	0***
Parity	-0.00935	0.01218	-0.768	0.4428
Sex	0.03599	0.03756	0.958	0.3381
AOM	0.0002	0.002705	0.073	0.9414
MES	0.08252	0.0401	2.058	0.0398*
GA	-0.00279	0.003666	-0.762	0.4463
BW	0.00001	0.000019	0.443	0.6577
BOOKING	0.09003	0.03977	2.264	0.0237*
YEAR	-0.00892	0.005832	-1.53	0.1262
Min	1Q	Median	3Q	Max
-1.2806	-0.2715	-0.1068	0.8099	1.0603

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Dispersion parameter for gaussian family taken to be 0.5747874)
 Null deviance: 961.79 on 1662 degrees of freedom
Residual deviance: 950.70 on 1654 degrees of freedom
AIC: 3809.5
Number of Fisher Scoring iterations: 2

The result of the multiple regression analysis presented in table 2 found that Mothers Education Status (MES) and Booking contributed significantly on the logistic model with p-values of 0.0398 and 0.0237 respectively which were found to be less than critical value of 0.05. Other factors such as Parity, Sex, Age of Mother (AOM), GA, Year, and Birth Weight (BW) were found to be insignificant on birth outcome in Anambra State.

Table 3: Confusion Matrix for Multiple Logistic Regression Analysis

		Predicted_BO		
		0	1	2
BO	0	1	379	0
	1	0	666	2
	2	1	615	0

E1= (379 + 0)/1664= 0.2278
E2= (0 +2)/1664= 0.0012
E3= (1 + 615)/1664 = 0.3702

Where,
E1 is the error rate for group 1, E2 is the error rate for group 2, and E3 is the error rate for group 3
Misclassification error rate = E1 + E2 + E3= 0.5992
The result presented in table 3 found that the misclassification error rate for birth outcome is 0.5992 (59.92%). This implies that 59.9% of birth outcomes were of the original cases was misclassified.

4.2. RESULT OF LINEAR DISCRIMINANT ANALYSIS

Table 4: Prior probabilities of groups

0	1	2
0.228503	0.40168	0.36981

The result obtained in table 4 showed that the prior probabilities of the groups were 0.228503, 0.40168 and 0.36981 for Alive, Neonatal death and Still birth respectively.

Table 5: Group means of the variables

	Parity	Sex	AOM	BOOKING	MES	GA	BW	YEAR
0	2.05	0.471053	31.66053	0.547368	0.542105	32.26053	2450.076	6.48947
1	1.757485	0.434132	30.13024	0.649701	0.685629	31.77844	2219.229	5.65719
2	1.910569	0.492683	31.26016	0.653659	0.648781	31.89268	2432.624	5.93659

The result obtained in table 5 shows the group means for the variables by the three birth outcomes Alive, Neonatal death and still birth.

Table 6: Result of Standardized Canonical Discriminant Function Coefficient

	Function	
	LD1	LD2
Parity	-0.16462	-0.01245
Sex	-0.24345	0.78639
AOM	-0.03995	0.03684
BOOKING	0.45908	1.06807
MES	0.95693	0.52763
GA	-0.02395	-0.02496
BW	-0.00047	0.00053
YEAR	-0.10931	-0.05216

The result obtained in table 6 showed the standardized coefficient for Maternal Education Status in the first function is greater in magnitude than the coefficients for the other variables with a coefficient of 0.95693 while the standardized coefficient for Booking Status in the second function is greater in magnitude than the coefficients for the other variables with a coefficient of 1.06807. Hence, Maternal Education Status and Bookings Status will have the greatest impact for first and second function respectively.

Table 7: Confusion Matrix for Linear Discriminant Analysis

		Predicted_BO		
		0	1	2
BO	0	24	192	164
	1	28	417	223
	2	33	347	236

E1= (192 + 164)/1664= 0.2139
E2= (28 +223)/1664= 0.1508
E3= (33 + 347)/1664 = 0.2284
Misclassification error rate = E1 + E2 + E3= 0.5931
The result presented in table 7 showed that the misclassification error rate for birth outcome is 0.5931 (59.31%). This implies that 59.3% of birth outcomes were of the original cases was misclassified.

4.3. RESULT OF QUADRATIC DISCRIMINANT ANALYSIS

Table 8: Prior probabilities of groups

0	1	2
0.228503	0.40168	0.36981

The result obtained in table 8 showed that the prior probabilities of the groups were 0.228503, 0.40168 and 0.36981 for Alive, Neonatal death and Still birth respectively.

Table 9: Group means of the variables

	Parity	Sex	AOM	BOOKING	MES	GA	BW	YEAR
0	2.05	0.471053	31.6605	0.547368	0.542105	32.2605	2450.08	6.48947
1	1.75749	0.434132	30.1302	0.649701	0.685629	31.7784	2219.23	5.65719
2	1.91057	0.492683	31.2602	0.653659	0.64878	31.8927	2432.62	5.93659

The result obtained in table 9 shows the group means for the variables by the three birth outcomes Alive, Neonatal death and still birth.

Table 10 : Confusion Matrix for Quadratic Discriminant Analysis

		Predicted_BO		
		0	1	2
BO	0	85	139	156
	1	82	385	201
	2	105	308	203

E1= (139 + 156)/1664= 0.1773
E2= (82 +201)/1664= 0.1701
E3= (105 + 308)/1664 = 0.2482
Misclassification error rate = E1 + E2 + E3= 0.5956

The result presented in table 10 revealed that the misclassification error rate for birth outcome is 0.5956 (59.56%). This implies that 59.6% of birth outcomes were of the original cases was misclassified.

5. CONCLUSION

This study employed multiple logistic regression, linear discriminant analysis and the quadratic discriminant analysis for estimating infant birth outcome and misclassification error rate of the birth outcomes in Anambra State. The birth outcomes of interest were the neonatal death, still birth and Alive. The result of the findings using the multiple logistic regression analysis showed that Mothers Education Status (MES) and Booking contributed significantly on the logistic model while factors of Parity, Sex, Age of Mother (AOM), GA, Year, and Birth Weight (BW) were found to be insignificant on birth outcome in Anambra State.

The misclassification error rate for birth outcome in Anambra State using the multiple regression approach is 0.5992 (59.92%). This indicated that 59.9% of birth outcomes were of the original cases was misclassified to estimate the birth outcome. Also, findings of the study equally showed that the prior probabilities of the groups for the linear and quadratic discriminant analysis were 0.228503, 0.40168 and 0.36981 for Alive, Neonatal death and Still birth respectively. Further findings revealed that the standardized coefficient for Maternal Education Status in the first linear function is greater in magnitude than the coefficients for the other variables while the standardized coefficient for Booking Status in the second linear function is greater in magnitude than the coefficients for the other variables. This result indicated that the Maternal Education Status and Bookings Status have the greatest impact for first and second linear function respectively. In addition, the result of the misclassification error rate for birth outcome using the linear discriminant analysis is 0.5931 (59.31%). The misclassification error rate for birth outcome using quadratic discriminant analysis is 0.5956 (59.56%). Based on the findings of this study, linear discriminant approach is the best alternative in estimating misclassification error rate of infant birth outcome followed by quadratic discriminant analysis and the least is multiple logistic regression .The findings clearly confirmed that the linear discriminant analysis is best with misclassification error rate of 59.31% which is in line with findings by Liu et al. (2016) who found that one-fifth of 365 neonatal deaths identified in the full birth history (FBH) survey were classified as stillbirths in the verbal/social autopsy (VASA) survey. The study recommends linear discriminant method for estimating the misclassification error rate of infant birth outcome in Anambra State and urgent need for design and implementation of policies, projects, and programmes that will give priority to essential child care thereby improving the quality of life of Nigerian child.

REFERENCES

[1] Adepoju, A. O. (2015). Differential Pattern in Child Mortality Rate in Rural Nigeria. Annual Research and Review in Biology, 7(5): 309-317.

[2] Adetoro, G. W. and Amoo, E. O. (2014). A Statistical Analysis of Child Mortality: Evidence from Nigeria. Journal of Demography and Social Statistics, Department of Demography and Social Statistics, ObafemiAwolowo University, Ile-Ife, Nigeria, 1: 110-120.

[3] Amzat, K. T. and Adeosun, S. A. (2014). On A Sequential Probit Model of Infant Mortality in Nigeria. International Journal of Mathematics and Statistics Invention (IJMSI), 2(3): 89-94

[4] Fisher, F.A. (1936). The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics, 7: 179-188.

[5] Hoffmann, J. (2003). Generalized linear models: An applied approach. Boston, MA: Allyn& Bacon.

[6] Jacdonmi, I., Suhainizam, M. S. and G. R. Jacdonmi, G. R. (2016). Breastfeeding, a child survival strategy against infant mortality in Nigeria. Current Science, 110(7): 1282-

[7] Liu, L., Kalter, H.D., Chu, Y., Kazmi, N., Koffi, A.K., Amouzou, A., Joos, O., Munos, M., and Black, R. E. (2016). Understanding Misclassification between Neonatal Deaths and Stillbirths: Empirical Evidence from Malawi. PLoS ONE, 11(12): 1-11.

[8] Richardson, K. E., Innocent, A. I., and Okereke, O. S. (2014). Relationship between Household Income and Child Mortality in Nigeria. American Journal of Life Sciences. Special Issue: Science, Society and Policy: Driving Towards Utopia or Dystopia, 2(6-4):1-12.

[9] Uddin, M., Hossain, M., and Ullah, M.O. (2009). Child Mortality in a Developing Country: A Statistical Analysis. Journal of Applied Quantitative Method, 4(3).

[10] Wolfgang, H., and Léopold, S. (2007). Applied Multivariate Statistical Analysis (2nd Ed.). Springer-Verlag Berlin Heidelberg New York, pages: 289-303.