FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 8 No. 3, June, 2024, pp 36 - 42
DOI: https://doi.org/10.33003/fjs-2024-0803-2472

AN APPLICATION OF POISSON REGRESSION MODEL ON ROAD ACCIDENTS DATA IN KADUNA STATE, NIGERIA

***[1]Sameer A. Saeed, [1]Yusuf Isah and [2]Yusuf Mustapha**

[1]Department of Mathematical Sciences, Federal University Dutsinma, Katsina State Nigeria.
[2]Department of Mathematical Sciences, Nigerian Defense Academy, Kaduna State Nigeria.

*Corresponding authors' email: Ssameer@fudutsinma.edu.ng

ABSTRACT
The Poisson Regression model was used to analyze the data on road traffic crashes of Kaduna State from year 2014 to 2017. The data was collected from the Kaduna State Command of the Federal Road Safety Corps (FRSC). The analysis was carried out using the R (MASS Package) software. The variables considered are; Number of Persons Involved in accidents, Season (months of the year), Number of Crashes and Causes of Accidents. The results from the model were analyzed using the Akaike Information Criterion (AIC) and goodness of fit. Result shows that the Poisson Regression Model has an AIC value of 1185.7. The scatter-plots are clustered and it has few outliers from the predicted line. This is due to the good value of the deviance and the parameters of the model are very much estimated. The result of the Poisson Regression Model aids in overcoming the over-dispersion problem, resulting in a better fit. Thus, indicating the existence of significant dispersion.

Keywords: Poisson, Model, Accidents, Akaike Information Criterion (AIC)

INTRODUCTION

In Nigeria today, barely a day goes by without the manifestation of a road traffic accident prominent to generally growing occurrence of injury and impermanence rates as well as financial cost to both society and the individual involved. Nigeria loses about 80billion Naira yearly to road accidents. Among all subjects that are involved in road traffic accidents in Nigeria, 29.1% suffer disability and 13.5% are unable to return to work (Labinjo, *et. al*, 2010). Nigeria is classified second highest in the rate of road accidents among 193 countries of the world (Agbonkhese, *et. al*., 2013). The (WHO, 2013) declared Nigeria the most insecure country in Africa with 33.7% deaths per 100,000 population every year. Rendering to their report, one in every four road accident deaths in Africa happens in Nigeria. The (WHO) analysis and the (FRSC, 2009) report of 5,693 fatal road accidents in 2009 leave no qualm about the risky condition on Nigerian roads. According to (Sheriff, 2009), Nigeria has the uppermost road accidents level as well as the major number of deaths per 10,000 vehicles. One may be absorbed to be certain of that the level of consciousness on the causes of road traffic accidents is very small amongst Nigerians.

In February 1988, the Federal Government of Nigeria recognized the Federal Road Safety Commission (FRSC). The Commission is accused with the following duties: making the highway safe for motorists and other road users, endorsing works and devices designed to eradicate accidents on the highways and recommending the federal and state governments covering the Federal Capital Territory Administration and significant governmental agencies on the sections where such works and devices are essential, and educating motorists and members of the public on the significance of punishment on the highway among others (FRSC., 2018). Regrettably, the once proud association has lost its focus, worthless the goodwill of the general public, and failed to meaningfully reduce death rate on our highways (Enahoro, 2018).

In 2017, the Kaduna State House of Assembly endorsed the law founding the Kaduna State Traffic Law and Enforcement Agency (KASTELEA). The new body will direct and control traffic on public roads in the state, examine vehicles, issues certificate of road merit, seize erring vehicles, and educate motorist on road safety procedures. The new body has also come short of its accountabilities as most of its focus is guided towards engendering revenues rather than the public safety. It is however shocking that in spite of the several bodies recognized to tackle road safety problematic, the degree of accident is still not significantly reduced. This dominant situation calls for scientific interference so as to realize the greatness of road accident, principally in Kaduna State**.**

In common, road traffic accidents lead to death and incapacity as well as commercial cost to both society and the individual involved. The reasons of road traffic accidents are not just human mistake or driver carelessness alone. Regrettably, Nigerian highways are debatably one of the worst and most dangerous in the world. According to (Elechi and Etawo, 1990), distress is the main reason for spare Room visits in Nigeria and road traffic accidents are accountable for the mainstream of deaths. Road traffic accident in Nigeria has not established the attention warranted so it lingers to be rising difficult all over the country. There is need to view road traffic accident as a dispute that needs pressing attention aimed at dropping the health, social and economic influences.

The Poisson Regression Model will be used to determine the better fit in handling the data of road traffic crashes within Kaduna state. The criterion for selection is the model which has a small AIC value, resulting in better estimation. This paper will look at the significant evidence in the Poisson Regression Model that, all the variables are statistically connected and affecting the number of persons involved. We will examine the plot of the Poisson model which will lie between some certain range and observe if it has smaller values of the Standard Deviation Residuals. We will also look at the chi-square test for independence that will indicate a certain P-Value which will be < 0.05. However, in spite of the development made over the years, emerging justifiable solutions to moderate road safety difficulties requires a deep considerate of the crack data features, and persuading factors. Accident data lean towards exhibit over dispersal, and yet, best accident data are modeled using Poisson Regression Model.

MATERIALS AND METHOD

The analysis of data on the number of people involved in road accident in Kaduna State and the comprehensive analysis of the data using statistical tools such as Poisson regression is to fit a model to the data set. This study is based on the data obtained from the Federal Road Safety Corps (FRSC), Kaduna State sector command. The method used for this study enlightened the theory behind the distributions and model for this application. The probability distributions of accident data and their possible regression model which may contain the Poisson distributions. It also delivered the account of the software package R (MASS Package) software used for the analysis and modeling.

Table 1: Kaduna State Road Accident Data

Season (months of the year)	Number of crashes	Persons involved	Number of causes
1	16	110	2
2	95	455	1
3	233	793	1
4	79	564	1
5	72	405	2
6	35	224	1
7	41	213	1
8	37	161	1
9	29	183	1
10	47	270	2
11	40	210	1
12	51	349	1
1	50	283	2
2	61	446	2
3	55	260	1
4	44	306	0
5	56	275	2
6	55	226	1
7	73	336	3
8	53	245	1
9	57	242	1
10	43	225	1
11	57	346	3
12	70	362	2
1	52	296	1
2	61	438	4
3	81	510	1
4	77	412	1
5	42	361	1
6	40	319	18
7	42	233	1
[8	63	312	4
9	83	430	0
10	45	243	18
11	48	281	0
12	92	465	1
1	105	463	18
2	80	268	18
3	10	574	18
4	101	498	18
5	90	563	18
6	72	333	18
7	51	353	3
8	82	425	2
9	56	328	18
10	59	295	18
11	62	275	18
12	64	347	18

Source: FRSC (2013) (Kaduna State sector command)

Monthly Distribution of number of Accidents that occurred yearly in Kaduna.

The Table 2 beneath displays the time in years for which accidents followed, it grants the total number of road accident yearly from 2014-2017and it also displays the number of road accidents in Kaduna State appears to be accumulating as years go by. Hence, we calculate the correlation matrix to verify the associate between the months and the number of accidents, and then we fixed the significant level at 0.05.

Table 2: Total Number of Road Accidents from 2014 to 2017

MONTHS	2014	2015	2016	2017	Total
JAN	16	50	52	105	223
FEB	95	61	61	80	297
MARCH	233	55	81	107	476
APRIL	79	44	77	101	301
MAY	72	56	42	90	260
JUNE	35	55	40	72	202
JULY	41	73	42	51	207
AUG	37	53	63	82	235
SEPT	29	57	83	56	225
OCT	47	43	45	59	194
NOV	40	57	48	62	207
DEC	51	70	92	64	277
TOTAL	775	674	726	929	3104

The Poisson Regression Model

The Poisson regression models are universal linear models with logarithm as the link function. In statistics, the Generalized Linear Model (GLM) is a stretchy overview of usual linear regression that allows response variables that have error distribution models other than a normal distribution. A generalized linear model is made up of a linear predictor.

$$\omega_i = \beta_0 + \beta_1 x_{i1} + \cdots \beta_k x_{ik} \quad (1)$$

And the two functions;

i. A Link function that defines how the mean, $E(y_i) = \lambda_i$, depends on the linear predictor

$$g(\lambda_i) = \omega_i \quad (2)$$

ii. A variance function that describes how the variance, $var(y_i)$depends on the mean

$$var(y_i) = \alpha\ var(\lambda) \quad (3)$$

Where the dispersion parameter α is a constant.

Assume $y_i \sim$ Poisson (λ_i) , then $E(y_i) = \lambda_i \ \ var(y_i) = \lambda_i$ (4)

Consequently, our variance function is $v(y_i) = (\lambda_i)$ (5)

And the link function must map from$(0, \infty) \rightarrow (-\infty, \infty)$. A natural choice is

$$g(\lambda_i) = log\ (\lambda_i) \quad (6)$$

The (GLM) generalizes linear regression by permitting the linear model to be connected to the response variable via a link function. Link function at this point is the function that links the linear model in a design matrix and the Poisson distribution function. Reflect a linear regression model given as;

$$y_i = \beta_0 + \beta_i x_i + \varepsilon_i \qquad i = 1,2,3 \ldots \quad (7)$$

If$x \in R^n$, is a vector of independent variables, then

$$Y = x\beta + \varepsilon \quad (8)$$

Where $x\ is\ an\ n \times (K+1)$ vector of independent variables, and a column of 1's $\beta\ is\ a\ (k{+}1)$ vector of regression parameters and $\boldsymbol{\varepsilon}$ signifies the measure of exposure. Hence,

$$E(Y|X) = x\beta \quad (9)$$

Recall that, for Generalized Linear Models, we use the link function to transform

Y: $G(Y) = log_e(Y)$ (10)

Therefore, this can be written more efficiently as

$$log_e E(Y|X) = x\beta \quad (11)$$

Thus, given a Poisson regression model with parameter $\beta$and an input vector x, the predicted mean of the related Poisson distribution is given by

$$E(Y|X) = e^{x\beta} \quad (12)$$

If y_i, are independent observations with consistent values x_i of the predicted variables, then β can be projected by maximum likelihood. The maximum likelihood evaluations lack a closed form countenance and must be originate by numerical methods. The probability surface for maximum likelihood Poisson regression is always convex, making Newton-Raphson or other gradient-based approaches suitable estimation techniques. Consequently, let y_i be the random variable, which takes non-negative values, i=1, 2… Where n is the number of observations. Since y_i follows a Poisson distribution, the probability mass function (PMF) is;

$$P(Y_i) = y_i \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \qquad y_i = 0, 1, 2 \ldots \quad (13)$$

Where $P(y)$the probability of number of individuals involved in accident and λ is the parameter that measures both the mean and the variance. The major hypothesis of Poisson model is

$$E(y_i) = var(y_i) = \lambda_i \quad (14)$$

Where the predicted mean of the Poisson distribution as given in (14) above quantified

$E(Y|X) = e^{x\beta} = \lambda_i$ (Same as the mean of the Poisson) $x$is the value of the descriptive variables and β = $(\beta_1, \beta_2, \ldots, \beta_k)$ are the unknown k-dimensional vector of regression parameters. The mean of the predicted Poisson distribution is given by $E(y)$ and variance of y_i as $var(y)$.The parameters λ can be estimated by maximum likelihood estimation method;

$$ln(\lambda) = \prod_{i=1}^{n} \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \quad (15)$$

The log-likelihood function is given by; $ln(\lambda) = \sum_{i=1}^{n}[-\lambda_i + y_i ln\lambda_i - lny_i!]$

The major assumption of the Poisson model is equation (14)

RESULT AND DISCUSSION

Correlation of number of accidents: 2014, 2015, 2016 and 2017

Table 3 displays the P-Value and the Pearson correlation value. The Pearson correlation between the number of accidents in 2014 and that of 2015 is -0.035, which indicates that there is a negative relationship. Also, between 2014 and 2016 is 0.367, which indicates a moderate positive

relationship between the years. And between 2014 and 2017 the Pearson correlation grew even stronger at 0.514, indicating a fairly positive relationship.
Similarly, the Pearson correlation between 2015 and 2016 is 0.103 indicating positive correlation, and between 2015 and 2017 is about -0.468 indicating a negative relationship. 2016 and 2017 is a positive relationship at about 0.177. It could also be confirmed from the table that the number of road accidents in Kaduna seems to be increasing as years go by except for 2014/2015 and 2015/2017 where we have a negative correlation, even at that, in-between the 2014/2015 and 2014/2016 is still a moderate positive correlation at about 0. 330. Next, we determine whether the correlation coefficient is significant by comparing the P-Value to our significant level at 0.05. Now α of 0.05 indicates the risk of concluding that a correlation exists when, actually, no correlation exist is 5%. If p-value ≤ α then the correlation is significant but if p-value > α then we cannot conclude that the correlation is different from 0.
From table 3, it shows that all the variables have P-Values greater than the significant level of 0.05. Hence, there is inconclusive evidence about the significance of the association between the months and the number of accidents.

Table 3: Correlation of the number of accidents

YEAR	2014	2015	2016	2017
2014	1	-0.037	0.367	0.514
		0.909	0.240	0.087
2015	-0.037	1	0.103	-0.468
	0.909		0.751	0.125
2016	0.367	0.103	1	0.177
	0.240	0.751		0.582
2017	0.514	-0.468	0.177	1
	0.087	0.125	0.582	

It is interesting to note from the graph of Figure 1 that the month with the highest frequency of road accident in Kaduna, is March as against the assumption of the public that December has the highest occurrence of road accidents. This may be due to the fact that most schools go on break during that period and people travel a lot during the festive period of Esther which occurs in the month of March. Other reasons can also be sort for that.

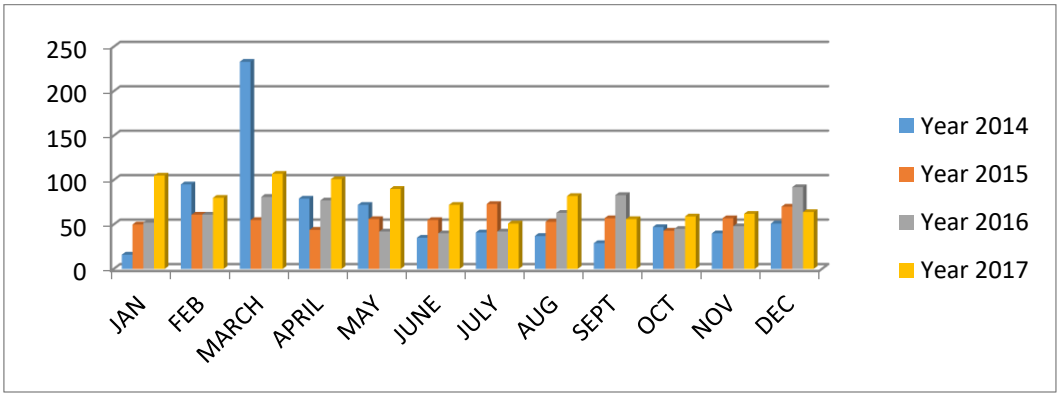

Figure 1: A bar graph illustrating the number of monthly road accident in Kaduna from 2014 to 2017

Four Year Distribution of the Total number of Persons Involved in Road Accidents in Kaduna
The Table 4 shows the sum of number of accidents (4 years) that occurred in a particular time (month) in years. It also presents the total number of persons involved in road accident in a particular month over the period of four years that is from 2014 - 2017. It can also be observed that as the number of accident increases, the number of persons involved in road accidents also increases. The 1st quarter (Jan to April) has the highest frequency of accident together with the highest number of persons involved to confirm the relationship between the number of accident and the number of persons involved, we compute chi-square test of independence. We set our significant level at 0.05.

Table 4: Four years Total Number of persons involved in Road Accidents monthly, from 2014 to 2017

MONTHS	TOTAL ACCIDENT	PERSON INVOLVED
JAN	223	1152
FEB	297	1607
MARCH	476	2137
APRIL	301	1780
MAY	260	1604
JUNE	202	1102
JULY	207	1135
AUG	235	1143
SEPT	225	1183
OCT	194	1033
NOV	207	1112
DEC	277	1523

Chi-Square Test of Independence of Table 1.3: Total Accident, Person Involved

The Chi-square parameter is estimated at Chi-Sq = 21.406, the degree of freedom for 12 months is DF = 11, and the probability of observing a sample statistic P-value is estimated at P-Value = 0.029. The result indicates the P-Value = 0.029 is < 0.05. Thus, we conclude that in each crashes people are involved, we find out that the number of persons involved is significantly different with the number of crashes and this is because some of the accidents are fatal, people are more involved than others. Similarly, people involved are not distributed over months. Some months there are more accidents involved and others not too much accident involved.

Table 5: Chi-Square Test of Independence

Months	Accident Total	Persons Involved	Total
1.Observed	223	1152	1375
Expected	217.59	1157.41	
Chi-sq value	0.135	0.025	
2. Observed	297	1607	1904
Expected	301.30	1602.70	
Chi-sq value	0.061	0.012	
3. Observed	476	2137	2613
Expected	413.50	2199.50	
Chi-sq value	9.448	1.776	
4. Observed	301	1780	2081
Expected	329.31	1751.69	
Chi-sq value	2.434	0.458	
5. Observed	260	1604	1864
Expected	294.97	1569.03	
Chi-sq value	4.146	0.779	
6. Observed	202	1102	1304
Expected	206.35	1097.65	
Chi-sq value	0.092	0.017	
7. Observed	207	1135	1342
Expected	212.37	1129.63	
Chi-sq value	0.136	0.025	
8. Observed	235	1143	1378
Expected	218.06	1159.94	
Chi-sq value	1.315	0.247	
9. Observed	225	1183	1408
Expected	222.81	1185.19	
Chi-sq value	0.022	0.004	
10. Observed	194	1033	1227
Expected	194.17	1032.83	
Chi-sq value	0.000	0.000	
11. Observed	207	1112	1319
Expected	208.73	1110.27	
Chi-sq value	0.014	0.003	
12. Observed	277	1523	1800
Expected	284.84	1515.16	
Chi-sq value	0.216	0.041	
Total	3104	16511	19615

Chi-Sq = 21.406, DF=11, P-Value = 0.029

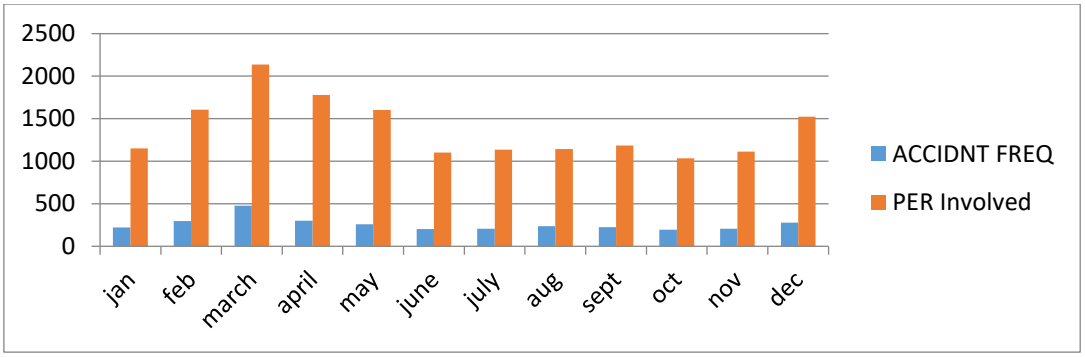

Figure 2: A bar graph illustrating the sum of number of persons involved in road accident in Kaduna monthly for four years (2014 to 2017)

Parameter Estimations of Poisson Regression Model from Kaduna State accident data

We explain the results we obtained from Poisson Regression Model after running the accident data. Table 6 and 7 present the parametric estimates for the Poisson regression model. These tables display the estimate of values of the parameters, standard error and the result from the AIC. It also shows that the independent variables i.e. Number of crashes, Season and Number of causes are significant to the Number of persons involved (dependent variable). Meaning as accidents are occurring and people are involved, there is much gap between the variables. The Poisson regression model has an AIC of 1185.7.

Table 6: Poisson Regression Model Parameter Estimation

Parameters	Estimation	Standard Error	z value	$Pr(>\|Z\|)$	significance
Intercept	5.4336676	0.0250195	217.177	< 2e-16	Significant
Number of Crashes	0.0064977	0.0001817	35.757	< 2e-16	Significant
Season (months of the year)	-0.0105924	0.0023835	-4.444	8.83e-06	Significant
Number of Causes	0.0042920	0.0010485	4.094	4.25e-05	Significant

Table 7: AIC Value of Poisson Regression Model Parameter Estimation

DISPERSION PARAMETER	1
DEVIANCE	84.23
AIC	1185.7

From Figure 3, it is clear that the scatter-plot is fairly clustered; it has few outliers from the predicted line. This is due to the good value of the deviance. That means the parameter of the model is partially estimated, hence it will lead to fair result, thus indicating the existence of average significant dispersion.

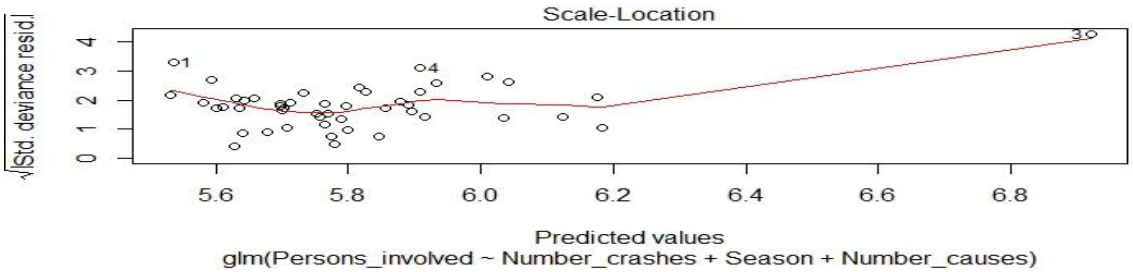

Figure 3: A plot of Poisson Regression Scale parameter

Figure 4 below is a normal quantile-quantile plot comparing independent standard deviation Residuals on the vertical axis to a standard normal theoretical quantile on the horizontal axis. The linearity of the points suggests that the data are normally distributed. However, the plot lies between -15 to +5 indicating the time it takes before an accident occurs is significantly long (i.e. there is a gap).

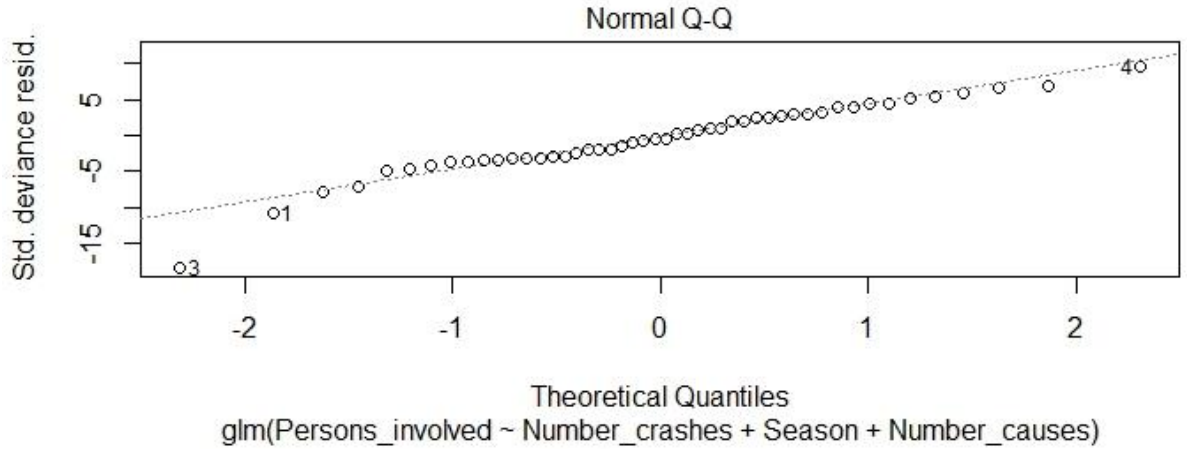

Figure 4: Normal Q-Q Plot of Poisson Regression model

CONCLUSION

The Poisson Regression Model was conducted to determine the better fit to use in handling the data of road traffic crashes within Kaduna state. The criterion for selection of the best model used, is the AIC. The best model is that with the smallest AIC value, resulting in better estimation. The study found that there is significant evidence in the Poisson Regression Model that, all the variables are statistically connected and affecting the number of persons involved. The plotted of the Poisson lies between the ranges 0.5 to 3.5 has smaller values of the Standard Deviation Residuals. The chi-square test for independence indicates the P-Value of 0.029 which is < 0.05. On the Basis of the AIC values from the Kaduna state data, the estimated AIC for Poisson Regression model is 1185.7.

However, the research also found that, the rate of accident and person involved in Kaduna state is significantly high because the Kaduna state highway is inter-linked with almost eight states, so naturally, the Kaduna state highway will be busy all the time. Hence, the data of road traffic crashes in Kaduna state is best modeled with Poisson Regression model.

REFERENCES

Agbonkhese, O., Yisa, G. L., Agbonkhese, E. G., Akanbi, D. O., Aka, E. O., & Mondigha, E. B. (2013). Road traffic accidents in Nigeria: Causes and preventive measures. *Civil and environmental research*, *3*(13), 90-99.

Elechi, E. N., & Etawo, S. U. (1990). Pilot study of injured patients seen in the University of Port Harcourt Teaching Hospital, Nigeria. *Injury*, *21*(4), 234-238.

Enahoro, E. (2018). FRSC and its Licensing Fiascos. *Retrieved from Daily Trust:* www.dailytrust.com.ng. March 26, 2018.

FRSC, (2016). FRSC end of year report. *Retrieved from* www.frsc.gov.ng. Jan 26, 2018.

FRSC, (2009). Traffic Digest. A Transport Digest Publication of *PRS Department, FRSC*.

Labinjo, M., Juillard, C., Kobusingye, O., & Hyder, A. A. (2010). Socioeconomic impact of road traffic injuries in West Africa: exploratory data from Nigeria. *Injury prevention*.

Sheriff, M. A. (2009). Traffic Education and Safety in Nigeria. *Nitours Journal*, 1(2).

World Health Organization. Violence, Injury Prevention, & World Health Organization. (2013). *Global status report on road safety 2013: supporting a decade of action*. World Health Organization.