

A Negative Binomial Regression on Road Accident Fatalities During COVID-19 Hit Era in Nigeria

Nwakuya Maureen Tobechukwu^{1,*}, Nwabueze Joy Chioma²

¹Department of Mathematics and Statistics, University of Port Harcourt, Port Harcourt, Nigeria

²Department of Statistics, Michael Okpara University of Agriculture, Umudike, Nigeria

Email address:

maureen.nwakuya@uniport.edu.ng (Nwakuya Maureen Tobechukwu), teeubueze@yahoo.com (Nwabueze Joy Chioma)

*Corresponding author

To cite this article:

Nwakuya Maureen Tobechukwu, Nwabueze Joy Chioma. A Negative Binomial Regression on Road Accident Fatalities During COVID-19 Hit Era in Nigeria. *International Journal of Statistical Distributions and Applications*. Vol. 8, No. 3, 2022, pp. 40-46.

doi: 10.11648/j.ijstd.20220803.11

Received: July 26, 2022; **Accepted:** August 10, 2022; **Published:** August 17, 2022

Abstract: The incidence of road fatality in the world is most alarming. It is noted that the number of deaths in Nigeria by COVID-19 is nothing compared to the deaths by road accidents. As at July 2022 the total number of deaths by COVID-19 is 3,144 while the number of deaths by road accident is 10116, which is about 69% higher than deaths due to COVID-19. This paper hence considered the road accident fatalities during the COVID-19 hit period (2019-2021) in all the states of Nigeria. Due to the count nature of the response variable (Number of deaths) Poisson regression and the negative binomial regression are considered in other to build a model for predicting number of deaths due to road accidents. Both methods were compared to determine the most appropriate. The results of the analysis show that the number of male fatality is more than double that of the female and children with a percentage of 70.6%. The data revealed that Akwa Ibom State and Kaduna State have the minimum and the maximum number of deaths with 47 and 1070 respectively. The Poisson regression and Negative Binomial (NB) regression were considered for analysis. The data showed evidence of over-dispersion rendering the Poisson regression inadequate for the analysis, leaving the negative binomial regression as an alternative. Further assessment of the two models based on the mean-variance relationship, goodness of fit test, AIC and BIC, identified the NB as a better model for the analysis. The model considered the number of people killed as the response variable while the covariates include; speed limit, car issues, dangerous driving, fatigue, street light, bad roads and general factors. Applying the NB regression produced an end result model with only dangerous driving as a significant factor in road accidents fatalities. This shows that dangerous driving is the major cause of deaths on Nigerian roads. The paper also noted that based on the available data, road accidents is a more threatening fatal hazard than COVID-19 with a total death toll of 3,144 for COVID-19 and 10,116 for Road accidents in the same period considered. The paper suggests that the government should put more weight in curbing road accidents by instructing the road safety commission to focus more in developing new strategies or improving on the old strategies for eliminating or reducing dangerous driving by drivers on Nigerian roads.

Keywords: Negative Binomial Regression, Poisson Regression, Over-dispersion, Goodness of Fit, Road Accident Fatalities

1. Introduction

The incidence of road fatality in the world is most alarming; World Health Organization (WHO) reveals that almost 1.2 million persons get killed from road accidents. Ameratunga S et al. [2] observed that deaths and wounds due to road accidents have effect on the society and economy at large and this is considerably on the increase. The population considered

in this study is all States in Nigeria with focus on the number of people that are killed by road accidents in all the 36 states and the Federal Capital Territory from 2019 to 2021. This period is we considered as COVID-19 hit era. Though COVID-19 is still on the loss in the world, this period considered in this work is the most hit period. Our interest was aroused by the realization that the number of deaths in Nigeria by COVID-19 in this period is nothing compared to the deaths by road accidents in the same period. As at July 2022 the total

number of deaths by COVID-19 is 3,144 while the number of deaths by road accident is 10,116, which tell us that deaths due to road accidents is about 69% higher than deaths due to COVID-19, in that same period. The paper aims at developing model to predictive deaths due to road accidents based on some factors. Data on deaths due to road accident constitutes a count data, hence it is considered to follow a discrete distribution. Since the response (number of killed) is a count data the Poisson regression and its modifications are considered in other to build this model.

Research works in Economics and Social Sciences often deal with count variables, hence modeling count data is a frequent task. In most cases the data usually display over-dispersion or in some cases evidence of too many number of zeros, hence limiting the use of the classical Poisson regression model. The various extension of the Poisson regression model has been proven to deal with this issue of over-dispersion and too many zeros. These extensions are models that belong to the family of generalized linear models and they include the use of sandwich covariance, the use of the quasi-Poisson model or the use the negative binomial (NB) regression [15, 16]. These mentioned models typically arrests over-dispersion quiet well, but they do not necessarily handle the issue of excess zeros. Lambert D [14] proposed the Zero-inflated Poisson (ZIP) regression model for handling count data with excess zeros. This triggered many interest, both in the econometrics and Statistics literature. Lambert D [14] proposed a mixture of models that brings together a count component and a point mass at zero. Some of the alternative models that can be considered for modeling count data are negative binomial and hurdle models. In this research work there are no non-zero counts therefore Zero-inflated Poisson (ZIP) regression was not considered. The study focused on the use of Poisson regression and Negative Binomial regression. The two models are investigated to ascertain best fit for the data. Our aim is to build model for predicting the incidents of deaths through road accidents based on some factors, identify significant factors that encourage road accidents and significant factors that intervene in road accidents deaths.

2. Linear Exponential Family

Generalized linear models (GLMs) explains the dependence of a scalar variable y_i ($i = 1, \dots, n$) on a vector of regressors x_i . In GLMs the conditional distribution of $y_i | x_i$ belongs to the linear exponential family with probability density function given as;

$$f(y; \vartheta, \lambda) = \exp\left(\frac{y\lambda - d(\lambda)}{\vartheta} + t(y\vartheta)\right) \quad (1)$$

where λ is the canonical parameter that depends on the s through a linear predictor and ϑ is a dispersion parameter that is often known. The functions $g(\cdot)$ and $t(\cdot)$ are known and determine which member of the family is used, e.g., the normal, binomial or Poisson distribution. Conditional mean and variance of the response y_i are given by'

$$E(y_i | x_i) = \mu_i = d'(\lambda_i) \text{ and } Var(y_i | x_i) = \vartheta d''(\lambda_i)$$

The dependence of the conditional mean $E[y_i | x_i] = \mu_i$ on the covariates x_i is specified via

$$g(u_i) = x_i^T \beta \quad (2)$$

where $g(\cdot)$ is a known link function and β is the vector of regression coefficients.

2.1. Poisson Regression

The Poisson regression model belongs to the family of generalized linear models (GLMs). In a Poisson regression analysis it is assumed that the response variable follows a Poisson distribution. Cameron, A. C. & Trivedi, P. K. [7] stated that Poisson regression models are suitable for situations in which the distribution of the event under study follows a Poisson distribution. The density function of the Poisson distribution is mathematically denoted as;

$$f(y) = \frac{\mu^Y \exp(-\mu)}{Y!}; Y = 0, 1, 2, \dots \quad (3)$$

Where the mean is $E(Y) = \mu$ and the variance $V(Y) = \mu$ and the dispersion index (DI) is given by

$$V(Y) / E(Y) = \frac{\mu}{\mu} = 1.$$

When the variance is greater than the mean over-dispersion occurs otherwise, it is called as under-dispersion. Over-dispersion has qualitatively similar consequences to the failure of the assumption of homoskedasticity in the linear regression model [6]. The Poisson regression hinges on the assumption that it is possible to model the logarithm of the expected outcome of the response variable by a linear combination of parameters whose values are not known. Poisson regression models the probability of non-negative integers (counts). Its model is given as;

$$Y_i = E(Y_i) + e_i, i = 0, 1, \dots \quad (4)$$

The i^{th} case mean response, is given by u_i which is assumed to be non-negative, and it is usually assumed to be a function of a set of predictor variables: X_1, X_2, \dots, X_{p-1} . Given that the distribution of the error terms e_i is a function of the distribution of the response y_i for a Poisson regression, we have the regression model given by:

$$\log(u_i) = X_i' \beta \quad (5)$$

where μ_i is the expected values of independent Poisson random variables, while X is a vector of the predictor variable and β is a vector of predictor effects. The regression coefficients $\beta_1, \beta_2, \dots, \beta_k$ are unknown parameters that will be estimated by the method of maximum likelihood. The variance-covariance matrix is given by; $(X' u_i X)^{-1}$. A multiplicative model of the mean is derived by exponentiating Equation (4) giving us;

$$u_i = \text{Exp}(X_i' \beta) \quad (6)$$

2.2. Negative Binomial Regression

The negative binomial (NB) regression model is based on the Poisson-gamma mixture distribution; it models the Poisson heterogeneity with a gamma distribution. Poisson-gamma model a mixture of two distributions as implied by the name and it was first derived by Greenwood and Yule [10]. The negative binomial can be parameterized in two distinct ways usually known as NB1 and NB2. This study focuses on NB2, hence is discussed below;

Let Y follow a Poisson distribution with parameter μ . Given below is the conditional distribution of the random variable Y ,

$$P(Y = \phi) = \frac{e^{-\mu} \mu^\phi}{\phi!} \quad \phi = 0, 1, 2, 3, \dots \quad (7)$$

Assuming ϕ follows a Gamma distribution with scale parameter ϕ and shape parameter β . Its probability density function is thus;

$$g(\phi) = \frac{\phi^\beta}{\Gamma(\beta)} \phi^{\beta-1} e^{-\phi}, \quad \phi > 0 \quad (8)$$

And the joint density of Y and ϕ is thus;

$$= \frac{\phi^\beta}{\Gamma(\beta)} \frac{\Gamma(y+\beta)}{(\phi+1)^{y+\beta}}$$

$$\therefore \text{the NB distribution is given by; } \frac{\Gamma(y+\beta)}{\Gamma(y+1)\Gamma(\beta)} \left(\frac{\phi}{\phi+1}\right)^y \left(\frac{1}{\phi+1}\right)^\beta, \beta > 0 \text{ and } \phi > 0 \quad (10)$$

3. Review of Literature

This section presents some past studies on count data analysis and road accidents. There are many research works on the travel behavior and choice of mode during the pandemic, researchers found that mode of travelling like walking and cycling significantly affected the means of transportation during the COVID-19 pandemic. Results also revealed that the use of public transport significantly went down. This can be found in the work of Bhaduri et al [4]. A study done in Japan by Inada et al [13] from January to May 2020 concluded that during the lock down violations of speed related issues and fatal road accidents were on the increase. While in the United Kingdom report from their provisional data showed a reduction on the volume of road traffic violations (measured in vehicle-miles) with a 16% downward trend in the number of road fatalities in 2020 in comparison to 2019. Evidence showed that the downward trend of road fatalities in 2020 was correlated to the decline in road traffic violations during COVID-19 era. Lambert D [14] proposed a comparative approach for handling count data by comparing five regression models on how they fitted their count data using the Akaike Information Criterion (AIC). Alexander K. M. et al. [1] investigated the performance of several count data models and to determine when they can be used.

Probability theory literature defines the Poisson process as process that counts the number of incidents that happens within a given time interval. The time interval that elapsed

$$P(Y=y|\phi = \theta)g(\phi) = \frac{e^{-\theta} \theta^y}{y!} \frac{\phi^\beta}{\Gamma(\beta)} \theta^{\beta-1} e^{-\phi} \quad (9)$$

Hence the unconditional distribution of Y from equation (9) above is obtained by summing out θ thus we have;

$$\begin{aligned} P(Y = y) &= \int_0^\infty P(Y = y|\phi = \theta)g(\phi) d\theta \\ &= \int_0^\infty \frac{e^{-\theta} \theta^y}{y!} \frac{\phi^\beta}{\Gamma(\beta)} \theta^{\beta-1} e^{-\phi} d\theta \\ &= \int_0^\infty \frac{\phi^\beta}{y! \Gamma(\beta)} \theta^{y-\beta-1} e^{-(\phi+1)} d\theta \\ &= \int_0^\infty \frac{\phi^\beta}{y! \Gamma(\beta)} \theta^{y-\beta-1} e^{-(\phi+1)} d\theta \\ &= \int_0^\infty \frac{\phi^\beta}{y! \Gamma(\beta)} \theta^{y-\beta-1} e^{-(\phi+1)} d\theta \\ &= \frac{\phi^\beta}{y! \Gamma(\beta)} \frac{\Gamma(y+\beta)}{(\phi+1)^{y+\beta}} \int_0^\infty \frac{(\phi+1)^{y+\beta}}{\Gamma(y+\beta)} \theta^{y-\beta-1} e^{-(\phi+1)} d\theta \end{aligned}$$

$$\text{Given that } \int_0^\infty \frac{(\phi+1)^{y+\beta}}{\Gamma(y+\beta)} \theta^{y-\beta-1} e^{-(\phi+1)} d\theta = 0$$

Hence we are left with;

between a pair of consecutive events is said to follow an exponential distribution with parameter lambda, were the individual inter-arrival times is assumed to be independent of other inter-arrival times. Alexander K. M. et al. [22] also observed that the Poisson process can also be used to model telephone calls at a call center, while Cannizzaro, F et al. [8], noted that the process is also very useful in modeling radioactive decay. According to Nixon [17], Poisson distribution is generally a go for option when modeling observed count data, with the assumption of equal mean and variance. Nevertheless, this assumption is rarely met in most applications due to the fact that data are usually over dispersed. In an over-dispersed data, the Poisson distribution's single parameter λ will not be able to fully give a detailed report on the event count hence causing the Poisson distribution to underestimate the dispersion of the observed counts. Nixon [17] noted that the Poisson regression is frequently opted for when modeling observed count data, with the assumption of equal mean and variance. Blackburn, M. L. [5] noted that the issue of over-dispersion is pertinent for a suitable choice between the Poisson and negative binomial regression model. Paternoster R et al. [19] with Paternoster R [18], as an alternative method recommended the utilization of negative binomial distribution in place of Poisson distribution in the presence of over-dispersion. Paternoster R [18] stated that over-dispersion is automatically corrected by the use of the negative binomial. He defined over-dispersion as a situation where the amount of variability in the model's fitted values is

much, and this is not consistent with Poisson regression assumption. Favero L. P. et al. [9] in their paper proposed a straightforward model selection approach that indicates the most suitable count regression model based on relevant data characteristics.

4. Methodology

4.1. Data Collection

The dataset for this project is retrieved from the National Bureau of Statistics bulletin of Nigeria. The individual records of dead victims from road accidents that occurred in all states of Nigeria including the Federal Capital territory of Nigeria from 2019 to 2021 with the factors that affected the incidents were obtained. The factors considered in this work include; number of individuals killed (killed) which is our response variable, the covariates include speed limit (SL), bad roads (BR), car issues (CI), fatigue (Fati.), and street light (SL), dangerous driving (DD) and general factors (GF). The models considered for fitting the data are the Poisson regression and the negative binomial regression. Below gives the description of model estimation. The analysis was done in R and estimation of NB applied the glm.nb function in R following procedure by Venables and Ripley [21].

$$\text{Estimate of mean number of deaths} = \text{Exp}(SL + CI + BR + DD + \text{Fati.} + SL + GF) \quad (13)$$

4.2.1. Estimation of Poisson Regression Model

Given the model in equation (3), the likelihood function is as follows:

$$L(\beta) = \frac{\prod_{i=1}^n (\mu(X\beta))^{Y_i} \exp(-\mu(X\beta))}{\prod_{i=1}^n Y_i!} \quad (14)$$

Once the functional form of $\mu(X\beta)$ is chosen, the maximization of (14) produces the maximum likelihood of the likelihood function:

$$\ln L(\beta) = \sum_{i=1}^n Y_i \ln[\mu(X\beta)] - \sum_{i=1}^n \mu(X\beta) - \sum_{i=1}^n \ln(Y_i!) \quad (15)$$

The maximum likelihood estimates $\beta_0, \beta_1, \dots, \beta_{p-1}$ are then derived through numerical search procedures. Iteratively reweighted least squares can again be used to obtain these estimates. After the maximum likelihood estimates are obtained, the model can be fitted. The goodness of fit of the model will be measured by the deviance. The deviance is given by;

$$D_p = \sum_{i=1}^n \left\{ y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right\} \quad (16)$$

If the observed value of the response is very close to the predicted mean of the response, the value of the deviance in equation (16) will be very small producing a good fitted model. The value of the residual deviance (RD) compared to the degrees of freedom (DF) will give an indication of presence or absence of over-dispersion. If the $RD < DF$ there is no over-dispersion and the Poisson model will be adequate but if the reverse is the case ($RD > DF$), the Quasi-Poisson and the Negative Binomial regression will be used.

4.2. Model Buildup

Let the response variable number of deaths (y_i) follow a Poisson distribution that is $y_i \sim P(\mu_i)$ with mean μ_i which is equal to the variance, where the mean depends on the covariates, we have;

$$\log(\mu_i) = X_i' \beta \quad (11)$$

where, μ_i is the mean response, X_i is the vector of covariates and β represents the model regression coefficient, it expresses the expected change in the log of the mean response per unit change in the covariate x_i . This simply means that one unit increase in the covariates x_i translates to an increase of β in the log of the mean response. A multiplicative model of the mean is derived by exponentiating Equation (10).

$$u_i = \text{Exp}(X_i' \beta) = \text{Exp}(\beta_0 + \sum_{r=1}^p x_{ir} \beta_r + e_i) \quad (12)$$

The exponentiated model yields an exponentiated regression coefficient $\exp\{\beta\}$ that expresses the multiplicative effect of the i^{th} covariate on the mean response. Hence a one unit increase in the covariate (X) multiplies the mean response by a factor $\exp\{\beta\}$. This log link is used because empirical observations have shown that with count data there is evidence of multiplicative effect of the covariates on the mean rather an additive effect. Our model is thus;

$$\text{Estimate of mean number of deaths} = \text{Exp}(SL + CI + BR + DD + \text{Fati.} + SL + GF) \quad (13)$$

The Akaike Information Criteria (AIC) as given by Hilbe J. M. [12];

$$AIC = \frac{2(L - (p+1))}{n} \quad (17)$$

and Bayesian Information Criteria as given by Hilbe J. M. [12];

$$BIC = D - (n - p - 1) \ln n \quad (18)$$

Where L is the log likelihood, p is the number of regression covariates, D is the deviance and n is the number of observations.

4.2.2. Estimation of Negative Binomial (NB) Regression Model

The regression coefficients of NB for this paper are estimated using the method of maximum likelihood, but if a person wants to use simulation-based estimation methods, you can check [11]. Let the distribution of the negative binomial be as in equation (9) and following the parameterization of

$\alpha = \frac{1}{\varphi}$ by [3, 20].

$$P(Y = y) = \frac{\Gamma(y+\alpha^{-1})}{y!\Gamma(\alpha^{-1})} \left\{ \frac{\alpha u}{1-\alpha u} \right\}^y (1 + \alpha u)^{-\frac{1}{\alpha}}, \alpha > 0, u > 0 \text{ \& } y = 0, 1, 2, \dots \quad (19)$$

It is regarded in a regression setting that $u_i = h^{-1}(v_i)$ where $h^{-1}(\cdot)$ represents the inverse of the link function and $v_i = x_i' \beta$ represents the linear predictor with x_i' being a vector of covariates and β a vector of regression coefficients.

Equation (19) can be rewritten as;

$$P(Y = y) = \frac{\Gamma(y+\alpha^{-1})}{y!\Gamma(\alpha^{-1})} \left\{ \frac{\alpha e^{x_i \beta}}{1 - \alpha e^{x_i \beta}} \right\}^y (1 + \alpha e^{x_i \beta})^{-\frac{1}{\alpha}} \quad (20)$$

The likelihood function as given by Cameron (2013);

$$L(\beta, \alpha) = \prod_{i=1}^n \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left\{ \frac{\alpha e^{x_i \beta}}{1 - \alpha e^{x_i \beta}} \right\}^{y_i} (1 + \alpha e^{x_i \beta})^{-\frac{1}{\alpha}} \quad (21)$$

And the log likelihood is given by;

$$\ln L(\beta, \varphi) = \sum_{i=1}^n (y_i \ln \alpha + y_i (x_i \beta) - (y_i + \alpha^{-1}) \ln(1 + \alpha e^{x_i \beta}) + \ln \Gamma(y_i + \alpha^{-1}) - \ln \Gamma(y_i + 1) - \ln \Gamma(\alpha^{-1})) \quad (22)$$

The values of β and φ that maximize $L(\beta, \alpha)$ will be the maximum likelihood estimates.

5. Results



Figure 1. Pie chart of deaths by States.

The plot in figure 1 above shows the proportion of deaths in all the states, the biggest proportion represents Kaduna State with the maximum deaths of 1070 and smallest chunk represents Akwa Ibom state with 47 deaths.

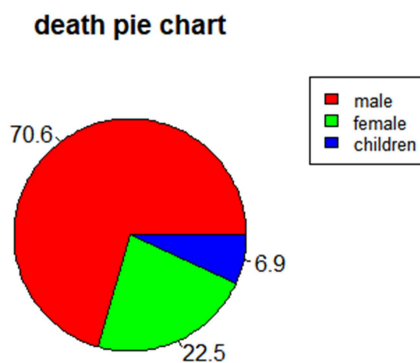


Figure 2. Pie chart according to groups.

Table 1. Group Distribution of Persons Killed.

Groups	killed	% killed
Male adult	7140	70.6
Female adult	2277	22.5
Children	699	6.9
TOTAL	10116	

Figure 2 and table 1 above displays the distribution of

deaths by groups namely; male female and children. The figure 2 and table 1 clearly shows that approximately 70% of people killed by road accidents are male adults, 23% are female adults and 7% are children. This is an indication that the male population is more prone to road accidents, hence resulting to most deaths in that group.

5.1. Poisson Regression Results

Null deviance: 5426.9 on 36 degrees of freedom

Residual deviance: 1858.8 on 29 degrees of freedom

ACI: 2139.7

BIC 2152.838

Table 2. Fitness of good test.

Deviance	Degrees of freedom	P-value
1858.751	29	0

The table 2 above, reveals that the Poisson regression model is not a good fit for the data, since the P-value is less than 0.05.

Table 3. Poisson Regression Model.

Parameter	Estimates	Std, error	z-value	P-value
Intercept	4.805	0.0208	231.08	<2.e16*
SL (β_1)	2.427e-05	1.6e-05	1.588	0.1124
CI (β_2)	7.983e-04	2.3e-04	3.521	0.0004*
BR (β_3)	7.100e-03	8.7e-04	8.158	3.41e-16*
DD (β_4)	2.268e-03	1.8e-04	12.628	<2e-16*
Fati. (β_5)	1.62e-02	8.3e-04	13.998	<2e-16*
SL (β_6)	-2.587e-03	1.5e-04	-17.76	<2e-16*
GF (β_7)	2.416e-03	3.5e-04	6.817	9.31e-12*

Significant at 0.05*.

The Poisson regression results in table 3 above shows that all the variables had a significant effect on the response except speed limit. The residual deviance obtained is higher than the degrees of freedom; hence there is presence of over-dispersion, in other words the Poisson regression

assumption has been violated and hence cannot be applied in the analysis.

5.2. Negative Binomial Regression Results

Null deviance: 98.149 on 36 degrees of freedom
Residual deviance: 38.252 on 29 degrees of freedom
AIC: 463.2164
BIC: 476.317
Dispersion parameter: 4.7115 (1.09)
2 x log-likelihood: -445.216

Table 4. Fitness of good test.

Deviance	Degrees of freedom	P-value
38.252	29	0.1168

The table 4 above reveals that the negative binomial regression model fits the data very well, since the P-value is greater than 0.05.

Table 6. Fitness of good test.

model	Deviance	DF	AIC	BIC	P-value (Goodness of fit)
Poisson	1858	29	2139.7	2152.838	0
Negative Binomial	38.252	29	463.22	476.317	0.1168

In the table 6 above based on the P-value of goodness of fit test, it shows that the negative binomial (NB) regression model is a good fit while Poisson regression model does not fit the data. The AIC of the NB is equally smaller than that of Poisson model, hence also an indication of a better fit. It can also be seen that the residual deviance of the Poisson regression model is higher than the degrees of freedom, which is an indication of over-dispersion, prompting the inadequacy of the Poisson model for this data analysis. Due to the goodness of fit of the NB and the presence of over-dispersion in the data, the NB is opted for as a better fit for the data.

Table 7. Exponentiated Estimates with Confidence Interval for NB regression Model.

Parameter	$e^{\text{Estimates}}$	UL	LL
Intercept (β_0)	107.136	0.9998	1.00025
SL (β_1)	1.0000	0.9965	1.00412
CI (β_2)	1.0004	0.9965	1.0041
BR (β_3)	1.0078	0.9955	1.0198
DD (β_4)	1.0041	1.0008	1.0076
Fati. (β_5)	1.00316	0.9871	1.0195
SL (β_6)	0.9984	0.9965	1.0006
GF (β_7)	1.0022	0.9975	1.0073

*UL means Upper limit.

* LL means lower limit.

Table 7 above presents the exponentiated estimates with the confidence intervals. The width of the confidence interval tells us that the estimates are very precise. The result reveals that the percent change in the incident rate of deaths due to road accidents has a 1% increase for every unit increase in the number of dangerous driving.

Table 5. Negative binomial Regression Model.

Variables	Estimates	Std error	Z-value	P-value
Intercept	4.674	0.1351	34.592	<2.e16*
SL (β_1)	5.07e-06	9.61e-05	0.053	0.9579
CI (β_2)	3.39e-04	1.77e-03	0.192	0.8479
BR (β_3)	7.72e-03	5.97e-03	1.294	0.1957
DD (β_4)	4.12e-03	1.49e-03	2.772	0.0056*
Fati. (β_5)	3.16e-03	7.17e-03	0.440	0.6596
SL (β_6)	-1.53e-03	1.14e-03	1.348	0.1776
GF (β_7)	2.22e-03	2.72e-03	0.817	0.4140

Significant at 0.05*.

The negative binomial regression results in table 5 above shows that all the variables had no significant effect on the response except for dangerous driving (DD). The AIC and the residual deviance were greatly reduced and the model has a dispersion parameter value of 4.7115. Using equation (12) and considering the significant variable below, the best model is thus;

$$(Expected\ mean\ deaths)u_i = e^{DD} \quad (23)$$

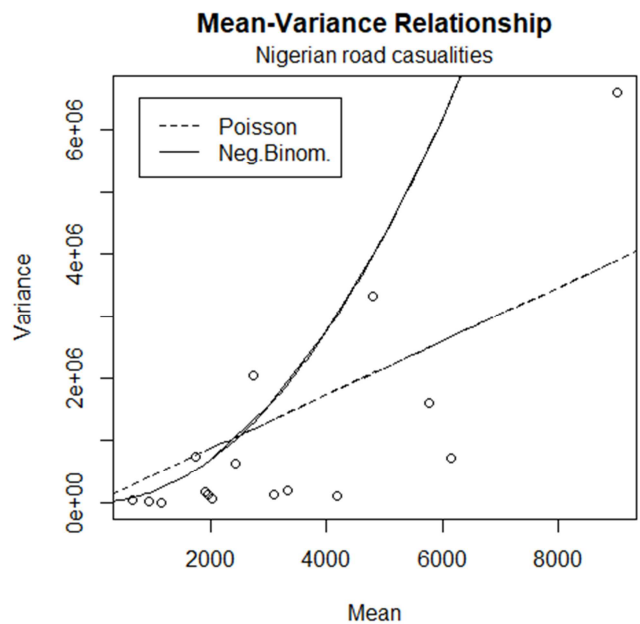


Figure 3. Mean-Variance Relation Plot.

The graph in figure 3 above plots the mean-variance relation where the overlaying curves represents the over-dispersed Poisson model, with variance $\phi\mu$, and the negative binomial model with variance $\mu(1+\mu\phi^2)$. The Poisson variance function has a good representation for majority of the data, but fails to capture the high variance. The negative binomial variance function is similar to it but, shows a quadratic form that can rise faster and hence capturing the high end.

6. Conclusion

In this paper we carried out analysis on deaths due to road accidents in Nigeria considering the COVID-19 hit period (2019-2020). The Poisson regression and Negative Binomial regression were considered. The result shows Akwa Ibom with 47 deaths as the state with the minimum number of deaths while the maximum of 1070 deaths was seen in Kaduna state. The study revealed that 70% of the deaths were male victims while approximately 23% were female and 7% were children. The study showed evidence of over-dispersion in the data making the Poisson regression inadequate for the analysis, leaving the negative binomial regression as an alternative. Further, an assessment of the Poisson regression and the negative binomial regression based on the mean-variance relationship, goodness of fit test, AIC and BIC revealed that Negative binomial model is a good fit for the data and provides a better description of the data. The end result model presented a parsimonious model with only dangerous driving as a significant factor in road accidents fatalities. This shows that in Nigeria dangerous driving is the major cause of deaths on Nigerian roads. The paper also noted that based on the available data from Nigeria, road accidents is a more threatening fatal hazard than COVID-19. The paper suggests that the government should put more weight in curbing road accidents and that the road safety commission should focus more in developing new strategies or improving on the old strategies for eliminating or reducing dangerous driving by drivers on Nigerian roads.

References

- [1] Alexander K. M., Oscar O. N., Anthony G. W. (2016). Statistical Models for Count Data. *Science Journal of Applied Mathematics and Statistics*. Vol. 4, No. 6, 2016, pp. 256-262.
- [2] Ameratunga S, Hajar M, Norton R. (2006). Road-traffic injuries: Confronting disparities to address a global-health problem. *Lancet*. 367: 1533-1540. Available. [http://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(06\)68654-6/fulltext](http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(06)68654-6/fulltext)
- [3] Anraku, K. and Yanagimoto T., (1990). Estimation for the negative binomial distribution based on the conditional likelihood. *Comm. Statist. Simul. Comput.*, 19: 771-786.
- [4] Bhaduri E., Manoj B. S., Zia W., Arkopal K. G, Charisma F. C. (2020). Modelling the effects of COVID-19 on travel mode choice behaviour in India. *Transportation Research Interdisciplinary Perspectives*; 8. <https://doi.org/10.1016/j.trip.2020.100273>.
- [5] Blackburn, M. L. (2015). The relative performance of Poisson and negative binomial regression estimators. *Oxford Bulletin of Economics and Statistics*: 77; 605-616.
- [6] Cameron A. C. and Trivedi P. K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of Applied Econometrics*, 1: pp 29-53.
- [7] Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data*, Cambridge university press.
- [8] Cannizzaro, F, Greco, G.; Rizzo, S. Sinagra, E. (1978) Result of the measurement carried in order to verify the validity of the Poisson-exponential distribution in radioactive decay event. *The international Journal of Applied Radiation and Isotopes*; 29 (11): 649.
- [9] Favero L. P., Paulo R., Paulo P. B., Hamilton L. C., Paulo M. F. C. (2021). *Count Data Regression Analysis: Concepts, Overdispersion Detection, Zero-inflation Identification, and Applications with R*. Practical Assessment, Research and Evaluation, 26: 13, 1-22.
- [10] Greenwood, M., and Yule, G. U. (1920). An Inquiry into the Nature of Frequency Distributions of Multiple Happenings, with Particular Reference to the Occurrence of Multiple Attacks of Disease or Repeated Accidents. *Journal of the Royal Statistical Society A*, 83: pp. 255-279.
- [11] Gourieroux, C. and A. Monfort (1997), *Simulation Based Econometric Methods*, Oxford: Oxford University Press.
- [12] Hilbe J. M. (2014). *Modeling Count Data*. Cambridge University Press, pp 126-161.
- [13] Inada, H., L. Ashraf and S. Campbell (2021). COVID-19 lockdown and fatal motor vehicle collisions due to speed-related traffic violations in Japan: A time-series study, *Injury Prevention*, Vol 27/1, pp. 98-100, <http://dx.doi.org/10.1136/injuryprev-2020-043947>.
- [14] Lambert D (1992). "Zero-inflated Poisson Regression, With an Application to Defects in Manufacturing." *Technometrics*, 34: pp 1-14.
- [15] McCullagh P, Nelder JA (1989). *Generalized Linear Models*. 2nd edition. Chapman & Hall, London.
- [16] Nelder JA, Wedderburn RWM (1972). "Generalized Linear Models." *Journal of the Royal Statistical Society A*, 135: 370-384.
- [17] Nixon D. C., (1991). Event count Models for Supreme Court dissents. *Political Methodology*; 4: 11-14.
- [18] Osgood, W. (2000). Poisson-based Regression Analysis of Aggregate crime Rates. *Journal of Quantitative Criminology*; 16: 21-43.
- [19] Paternoster R, Brame R. (1997). Multiple routes to delinquency; A test of developmental and general theories of crime. *Journal of criminology*; 35: 45-84.
- [20] Piegorsch, W. W., (1990). Maximum likelihood estimation for the negative binomial dispersion parameter. *Biometrics*, 46: 863-867.
- [21] Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth edition. Springer.
- [22] Willkomm, D.; Machiraju, S. Bolot, J. Wolisz, A. (2009). Primary user behaviour in cellular network and implications for dynamic spectrum access. *IEEE Communication magazine* 47 (3); 88.