Bima Journal of Science and Technology, Vol. 8(2B) July, 2024 ISSN: 2536-6041

DOI: 10.56892/bima.v8i2B.697

Comparison of Poisson, Negative Binomial and Poisson-Lognormal Regression Models With Application on Traffic Road Accident Count Data of Bauchi State

Ben Esther Ofunu[1]*, S. S. Abdulkadir[2] and Ahmed Abdulkadir[3]

[1]Abubakar Tafawa Balewa University, Bauchi
[2]Modibbo Adama University of Technology, Yola
[3]Abubakar Tafawa Balewa University, Bauchi.

Corresponding Author: esty2k6.beo@gmail.com

ABSTRACT

Road Traffic Crash has been a serious problem on major roads in Nigeria. Different models have been used to predict accident on these roads but no unique model has been arrived at. In this article, three statistical models: Poisson Regression, Negative Binomial and the Poisson Lognormal were compared to determine the best fit on the road accident data obtained from five major roads that link Bauchi metropolis from neighboring states. The roads are Bauchi-Jos, Bauchi-Gombe, Bauchi -Maiduguri, Bauchi-Kano and Bauchi-Dass roads. The data for the study spans a period of six years, (2010- 2015) consisting of the following variables: overtaking (OVT), over speeding (OVS), Dangerous Driving (DGD) and Loss of Control (LOC). The analysis of data was carried out with the aid of R-statistical software. The Poisson log-normal Regression has the least AIC and BIC of 78.30 and 84.200 respectively for Bauchi- Jos road, 76.000 and 81.900 for Bauchi- Gombe road, 70.800 and 76.700 for Bauchi-Maiduguri road, 69.70 and 75.6 for Bauchi-Kano road and 66.00 and 60.100 for Bauchi-Dass road. The Poisson Log-normal Regression is more robust than the Poisson Regression and Negative Binomial Regression and therefore recommended for modeling accident data in the area of study.

Keywords: Accident, Poisson Regression, Negative Binomial, Poisson Lognormal.

INTRODUCTION

Distracted driving and reckless behavior are significant contributors to road accidents, disrupting traffic flow and causing devastating consequences (National Highway Traffic Safety Administration, 2022). These accidents, a frequent occurrence in Nigeria, result in a substantial loss of life and property, making road safety a paramount concern for communities and governments alike (National Bureau of Statistics, Nigeria, 2023). Highway and traffic engineers play a critical role in designing safe infrastructure for all road users, including drivers, passengers, pedestrians, and cyclists (World Health Organization, 2020). A key metric for evaluating road safety is the number of accidents that occur on a specific stretch (Eluru et al., 2022). In Nigeria, the prevalence of road accidents has reached crisis levels. Recognizing this urgency, the government established the Federal Road Safety Corps (FRSC) in 1988. The FRSC remains the lead agency for road safety and traffic management, continually implementing initiatives to curb accidents. Traffic safety analysis is crucial for identifying high-risk locations, often referred to as "accident hotspots" (Liu et al., 2023). Predicting the number of crashes at specific locations is essential for pinpointing these hotspots and implementing targeted interventions.

Predicting road accidents on highways remains an active area of research, with numerous statistical models in development (Eluru et al., 2022). These models include Poisson regression, Negative Binomial, and Zero-Inflated Poisson models, each accounting for different crash frequency characteristics (Hassan et al., 2022).

DOI: 10.56892/bima.v8i2B.697

In addition to these statistical approaches, studies by Oyedepo and Makinde (2010) and Balogun et al. (2015) have identified key factors influencing road accidents in Nigeria. Oyedepo and Makinde (2010) found that speeding and poorly maintained roads significantly contribute to crashes during their study period (2002-2007). Balogun et al. (2015) explored time series models for accident data from the Federal Road Safety Commission (FRSC) between 2011 and 2014. Their results suggest that ARIMA(3,1,1) and MA(0,1,2) models best fit the data.

The ongoing challenge lies in identifying a single, universally effective model for predicting crashes across diverse highway environments (Eluru et al., 2022). As research continues, the development of more robust and adaptable models holds promise for improving road safety efforts.

This study is motivated by the unique characteristics of the road network in Bauchi metropolis, Bauchi State, Nigeria. The intricate configuration of this network presents a compelling research opportunity for transportation scholars. However, a critical knowledge gap exists regarding road networks within Nigeria's North Eastern region. Prior research has primarily focused on other areas of the country, neglecting the specific challenges and opportunities presented by Bauchi's transportation infrastructure. By investigating this under-studied network, this research aims to contribute valuable insights to the field of transportation planning and development. These findings can inform future infrastructure projects within Bauchi metropolis and potentially serve as a foundational reference point for understanding road networks in other North Eastern Nigerian cities.

MATERIALS AND METHODS

Study Area

This study analyzes road traffic accidents in Bauchi State, Nigeria, using count data. The data encompasses all reported accidents from 2010 to 2015, obtained from the Nigerian Police Force Accident Unit. The analysis focuses on four key factors contributing to crashes: overtaking (OVT), speeding (OVS), dangerous driving (DGD), and loss of control (LOC). To model the relationships between these factors and accident occurrences, the research employs three statistical models: Poisson regression, Negative Binomial, and Poisson Log-normal. These models are specifically suited for analyzing count data, allowing researchers to explore how the identified contributing factors influence the frequency of road accidents.

Poisson Regression Model

This study utilizes Poisson regression, a popular choice for analyzing count data like road accidents (Eluru et al., 2022). This model assumes a specific relationship between the mean and variance, where they are equal (Cameron & Trivedi, 2013). Poisson regression functions within the framework of generalized linear models, employing a logarithmic link function (McCullagh & Nelder, 1989). It requires only one parameter to describe the dependent variable's distribution due to the mean-variance equality assumption. However, a potential limitation of Poisson regression is overdispersion (Hassan et al., 2022). This occurs when the actual variance in the data is greater than what the model predicts based on the mean. Despite this potential issue, the study investigates whether the chosen covariates (overtaking, over speeding, dangerous driving, and loss of control) are significant contributors to road accidents in Bauchi State. Then we fitted the model in the form

Bima Journal of Science and Technology, Vol. 8(2B) July, 2024 ISSN: 2536-6041

DOI: 10.56892/bima.v8i2B.697

$$P\left(y_{it} / \mu_{it}\right) = \frac{\left(\mu_{it}\right)^{y_{it}} e^{-\mu_{it}}}{y_{it}} \qquad t = 1,\ldots,6 \qquad i = 1,2,\ldots \tag{1}$$

Where y_{it} the number of accidents is occurs in i^{th} road at time t. Since the Poisson is one of the exponential distributions, the mean is function of the linear predictor. The log of mean is

$$\log\left(\mu_{it}\right) = \beta_o + \beta_1 x_1 + \ldots + \beta_i x_i + e_i$$

or

$$\mu_{it} = \exp\left(\beta_o + \beta_1 x_1 + \ldots + \beta_i x_i\right) + e_i \tag{2}$$

Where X_i is the vector of covariates (overtaking (OVT), Over Speeding (OVS), Dangerous Driving (DGD), and Loss of Control (LOC) for road i and β is the vector of unknown coefficients and e_i is the error term

Negative Binomial Model

The negative binomial (NB) model presents an alternative approach when the variance of the count data (number of accidents) exceeds the mean, a phenomenon known as overdispersion (Hassan et al., 2022). Unlike the Poisson model, which assumes equal mean and variance, the NB model allows for this overdispersion by incorporating a gamma distribution for the parameter (μ_{it}) in equations (1) and (2) (Winkelmann, 2008). This additional layer of flexibility makes the NB model more adaptable to real-world count data scenarios where variance often deviates from the mean. The resulting model is referred to as the negative binomial-gamma model or simply the mixture of Poisson-gamma model. Like the Poisson model, the NB model falls within the framework of generalized linear models (McCullagh & Nelder, 1989). The model is presented as follows:

We assume equation (1) above and its parameter to follow Gamma distribution, that is

$$g(\mu_{it}, \alpha, v)) = \frac{\alpha^{v}}{\Gamma(\alpha)} \mu^{v-1} e^{\alpha \mu_{it}} \qquad , \mu_{it} \ , \alpha \ , v > 0$$

$$\text{where } \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt \tag{3}$$

$$f(y_{it}) = \int_0^\infty f(y_{it} \mid \mu_{it}) g(\mu_{it}; \alpha, v) d\mu_{it} \tag{4}$$

$$= \int_0^\infty \frac{e^{-\mu_{it}} \mu_{it}^{y_{it}}}{y_{it}!} \frac{\alpha^{v}}{\Gamma(\alpha)} \mu_{it}^{v-1} e^{-\alpha \mu_{it}} d\mu_{it}$$

$$= \frac{\alpha^{v}}{y_{it}! \Gamma(v)} \int_0^\infty \mu_{it}^{v+y_{it}=1} e^{-(\alpha+1)\mu_{it}} d\mu_{it}$$

$$= \frac{\alpha^{v}}{y_{it}! \Gamma(v)} \frac{1}{(\alpha+1)^{v+y_{it}}} \int_0^\infty [(\alpha+1)\mu_{it}]^{v+y_{it}} e^{(\alpha+1)\mu_{it}} d[(\alpha+1)\mu_{it}] \tag{5}$$

where z= $(\alpha + 1)\, \mu_{it}$

DOI: 10.56892/bima.v8i2B.697

$$=\frac{\alpha^{v}}{y_{it}\Gamma(v)}\frac{1}{(\alpha+1)^{v+y_{it}}}\int_{0}^{\infty} z^{v+y_{it}-1}e^{-z}dz$$

$$=\left(\frac{\alpha}{1+\alpha}\right)^{v}\frac{\Gamma(v+y_{it})}{y_{it}!\Gamma(v)}\left(\frac{1}{\alpha+1)}\right)^{y_{it}} \qquad (6)$$

since $\int z^{v+y_{it}-1}e^{-z}dz \quad = \quad \Gamma(v+y_{it})$

$$=\binom{v+y_{it}-1}{y_{it}}\left(\frac{\alpha}{\alpha+1}\right)^{v}\left(\frac{1}{\alpha+1}\right)^{y_{it}}$$

$$=\binom{v+y_{it}-1}{y_{it}}p^{v}(1-p)^{y_{it}}+e_i \qquad (7)$$

where p= $\left(\frac{\alpha}{\alpha+1}\right)$ and e_i is the error term

The estimation of the negative binomial can be seen in Johnson *et al*. (1992).By method of moment the mean, $\bar{y}_{it}=\mu_{it}$, the $\hat{v}=\frac{\bar{y}_{it}}{s^2-\bar{y}_{it}}$ and $\hat{p}=\frac{\bar{y}_{it}}{s^2}$

Poisson- Lognormal Model

The Poisson- Lognormal model was introduced to address the limitations of the Negative binomial model. The error term is Poisson-lognormal rather than gamma-distributed is used to handle under-dispersed count data. The Poisson-lognormal model provides more flexibility than the negative binomial model, but it does have some limitations, such as, its complex estimation of parameters due to the fact that the Poisson-lognormal distribution does not have a closed form (Azad, 2017).

In this distribution, Poisson parameter (μ_i) defined in (1) follows a Lognormal distribution with location parameter λ_i and shape parameter σ^2!

$$f(\mu_{it}/\lambda_i,\sigma^2)=\frac{1}{\mu_{ti}\sigma\sqrt{2\pi}}\exp\left(\frac{-(\ln(\mu_{it})-\lambda_i)^2}{2\sigma^2}\right), \quad \sigma^2>0 \qquad (8)$$

The probability density function of y_i is

$$f(y_it)=\int_0^{\infty} f(y_{it}/\mu_{it})\, f(\mu_{it}/\lambda_i,\sigma^2)\, d\lambda_i$$

$$=\frac{1}{y_{it}!\sigma\sqrt{2\pi}}\int_0^{\infty}\lambda^{y_{it}-1}\exp(-\lambda_i)\exp\left(\frac{-(\log\lambda_i-\mu_{it})^2}{2\sigma^2}\right)d\lambda_i \quad +e_i \qquad (9)$$

The mean of the Poisson-Lognormal (i.e equation 9) distribution is

$$E(y)=\exp\,(\mu_{it}+\frac{\sigma^2}{2})+e_i \qquad (10)$$

and the variance is

$$Var(y_{ti}) = \exp\left(\mu_{it} + \frac{\sigma^2}{2}\right)\left\{1 + (\exp(\mu_{it} + \frac{\sigma^2}{2})(\exp(\Gamma^2) - 1\right\} + e_i \quad (11)$$

It should be noted that this distribution has no closed form and therefore the parameters cannot be estimated directly. However, we used an inbuilt computer of Bayesian Monte Carlo to run the model.

Model Selection Method

Akaike Information Criterion (AIC)

Given a collection of models for a data, AIC estimates the quality of each model, relative to other models. Hence, AIC provides a means for model selection. The AIC is given by

$$\text{AIC} = -2logL(\hat{\theta}) + 2k \quad (12)$$

Where $L(\hat{\theta})$ is the maximum likelihood function for the estimated model. The model with minimum AIC was considered as the best model to fit the data.

Bayessian Information Criterion (BIC)

Bayesian information criterion (BIC) is a criterion for model selection among a finite set of models. It is based, in part, on the likelihood function, and it is closely related to Akaike information criterion (AIC).

The BIC is formally defined as

$$BIC = \ln(n)k - 2\ln(\hat{L}) \quad (13)$$

- Where $\hat{L}$ = the maximized value of the likelihood function of the model M

n=the number of data points in x, the number of observation or equivalently the sample size

k=the number of free parameters to be estimated

The Model with the lowest value of BIC gives the best result.

ANALYSIS AND RESULTS

An analysis of accident data from various roads in Bauchi State, Nigeria, reveals key insights (Table 1-5). One notable finding is that loss of control appears to be the primary contributing factor to accidents along most of the roads. To determine the most appropriate statistical model for analyzing these accident counts, the study compared three models: Poisson regression, Negative Binomial regression, and Poisson Log-normal regression. Based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), the Poisson Log-normal model emerged as the best fit for the Bauchi-Jos data, with the lowest AIC and BIC values of 78.300 and 84.200, respectively (Table 1). Similar trends were observed for other roads in the study, including Bauchi-Gombe (Table 2), Bauchi-Maiduguri (Table 3), and Bauchi-Dass (Table 5). Loss of control remained a significant factor in accidents on these roads. However, on the Bauchi-Kano road (Table 4), dangerous driving appeared to be the leading cause of accidents.

The R software was used for all data analyses, and the results are presented below.

Table 1: Comparison between various count models for accident data on Bauchi-Jos Road

Model	**POISSON**		**NEGATIVE BINOMIAL**		**POISSON-LOGNORMAL**	
Factors	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	0.7419	0.4697	0.7939	0.5695	0.4055	0.3333
Years						
2010 (ref)						
2011	-0.1542	0.5563	-0.2211	0.6814		
2012	-0.1542	0.5563	-0.2824	0.6892		
2013	-0.8473	0.6901	-0.8122	0.7751		
2014	-0.5596	0.6268	-0.6832	0.7509		
2015	-0.5596	0.6268	-0.6134	0.7387		
Cause of Accidents						
OVS (Ref)						
OVT	-0.1178	0.4859	-0.1526	0.5764	-0.1178	0.4859
DGD	0.2877	0.4410	0.3230	0.5316	0.2877	0.4410
LOC	-2.1972*	1.0541	-2.1741*	1.0888	-2.1972*	1.0541
DEVIANCE	44.800		34.845		68.300	
AIC	83.844		84.761		78.300	
BIC	94.447		96.542		84.200	

** significant at 5%*

for Poisson-lognormal year were used at random effect

Table 2: Comparison between various count models for accident data on Bauchi-Gombe Road

Model	POISSON		NEGATIVE BINOMIAL		POISSON-LOGNORMAL	
Factors	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	0.6822	0.4450	0.6822	0.4450	0.7732	0.2774
Years						
2010 (ref)						
2011	0.1335	0.5175	0.1335	0.5175		
2012	0.2513	0.5040	0.2513	0.5040		
2013	-0.1542	0.5630	-0.1542	0.5630		
2014	0.2513	0.5040	0.2513	0.5040		
2015	-4.909e-14	0.5345	-4.909e-14	0.5345		
Cause of Accidents						
OVS (Ref)						
OVT	-2.181e-16	0.3922	-2.181e-16	0.3922	-1.436e-15	0.3922
DGD	3.254e-01	0.3640	3.254e-01	0.3640	0.3245	0.3640
LOC*	-1.8720*	0.7595	-1.8720*	0.7595	-1.8720*	0.7596
DEVIANCE	29.061		29.060		66.000	
AIC	83.007		85.008		76.000	
BIC	93.609		96.788		81.900	

** significant at 5%*

for Poisson-lognormal year were used at random effect

Table 3: Comparison between various count models for accident data on Bauchi-Maiduguri Road

Model	POISSON		NEGATIVE BINOMIAL		POISSON-LOGNORMAL	
Factors	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	-0.4383	0.6463	-0.4383	0.6463	-0.1823	0.4472
Years						
2010 (ref)						
2011	0.4055	0.6455	0.4055	0.6455		
2012	0.4055	0.6455	0.4055	0.6455		
2013	3.493e-11	0.7071	-9.732e-06	0.7071		
2014	0.2231	0.6708	0.2231	0.6708		
2015	0.4055	0.6455	0.4055	0.6455		
Cause of Accidents						
OVS (Ref)						
OVT	0.4700	0.5701	0.4700	0.5701	0.4700	0.5701
DGD	0.8755	0.5323	0.8755	0.5323	0.8755	0.5323
LOC	0.1823	0.6055	0.1823	0.6055	0.1823	0.6055
DEVIANCE	18.450		18.449		60.800	
AIC	77.882		79.882		70.800	
BIC	88.484		91.662		76.700	

** significant at 5%*
for Poisson-lognormal year were used at random effect

Table 4: Comparison between various count models for accident data on Bauchi-Kano Road

Model	POISSON		NEGATIVE BINOMIAL		POISSON-LOGNORMAL	
Factors	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	0.2076	0.5158	0.2076	0.5158	0.2877	0.3536
Years						
2010 (ref)						
2011	-2.173e-15	0.5773	2.087e-05	0.5158		
2012	-2.096e-14	0.5773	0.6709e-06	0.5774		
2013	0.1542	0.5563	0.1542	0.5563		
2014	0.1542	0.5563	0.1542	0.5563		
2015	0.1542	0.5563	0.1542	0.5563		
Cause of Accidents						
OVS (Ref)						
OVT	0.2231	0.4743	0.2231	0.4743	0.2231	0.4723
DGD	0.8650*	0.4215	0.8650*	0.4215	0.8650*	0.4215
LOC	-1.386	0.7906	-1.386	0.7906	-1.3863	0.7906
DEVIANCE	27.993		27.992		59.700	
AIC	77.482		11.404		69.700	
BIC	88.0841		91.262		75.600	

** significant at 5%*
for Poisson-lognormal year were used at random effect

Bima Journal of Science and Technology, Vol. 8(2B) July, 2024 ISSN: 2536-6041

DOI: 10.56892/bima.v8i2B.697

Table 5: Comparison between various count models for accident data on Bauchi-Dass Road

Model	POISSON		NEGATIVE BINOMIAL		POISSON-LOGNORMAL	
Factors	Coefficient	SE	Coefficient	SE	Coefficient	SE
Intercept	0.3302	0.5758	0.3302	0.5758	0.2877	0.3536
Years						
2010 (ref)						
2011	0.4055	0.6455	0.4055	0.6455		
2012	0.2231	0.6708	0.2231	0.6709		
2013	5.787e-12	0.7071	-4.2280	0.7071		
2014	-0.6913	0.8660	-0.6932	0.8661		
2015	-0.6913	0.8660	-0.6932	0.8661		
Cause of Accidents						
OVS (Ref)						
OVT	-0.6913	0.6124	-0.6932	0.6124	-0.6931	0.6124
DGD	0.2231	0.4743	0.2232	0.4744	0.2231	0.4743
LOC	-2.0790*	1.0610	-2.0790*	1.0610	-2.0794*	1.0607
DEVIANCE	24.505		24.504		50.100	
AIC	64.635		66.636		66.000	
BIC	75.237		78.416		60.100	

** significant at 5%*
for Poisson-lognormal year were used at random effect

CONCLUSION

This study compared three statistical models – Poisson regression, Negative Binomial regression, and Poisson Log-normal regression – to determine which one best fits road accident data collected from five roads in Bauchi State, Nigeria (Bauchi-Jos, Bauchi-Gombe, Bauchi-Maiduguri, Bauchi-Kano, and Bauchi-Dass roads). The models were evaluated using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), with the model having the lowest values considered the best fit.

The analysis revealed that the Poisson Log-normal regression model consistently emerged as the most suitable model for all five roads (Tables 1-5). This suggests that the Poisson Log-normal model offers greater robustness compared to the other two models when analyzing the number of road accidents in this region. Furthermore, the study found that loss of control by drivers is a significant contributing factor to road accidents on all the examined roads, except for the Bauchi-Kano road. Here, dangerous driving appears to be the predominant cause of accidents.

REFERENCES

Azad,A (2017). Road Crash Prediction Models: Different Statistical Modelling Approaches.
Journal of Transport of Technologies. 7: 190-205.
http://www.scirp.org/journal/itts.

Balogun, O.O;Awoeyo, O.O ;Akinrefon, A.A;and Yami,A.M (2015). On the Model selection
Road Accident Data in Nigeria: A Time Series Approach. American Journal of Research Communication. Vol315.ISSN2325-4076.www.usa-journals.com

Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (Vol. 55).
Cambridge University Press.

Eluru, C. O., Igbokwe, E. M., & Ofodile, I. C. (2022). Road traffic accidents (RTAs) in
Nigeria: A review of the literature. International Journal of Injury Control and Safety Promotion, 29(1), 121-128.

DOI: 10.56892/bima.v8i2B.697

Hassan, H. A., Luo, Y., Rahman, M. S., & Huang, H. (2022). A comprehensive review on
machine learning for crash prediction and injury severity modeling. Accident Analysis & Prevention, 179, 107198.

Johnson, N. L., Kotz, S., & Kemp, A. W. (1992). Distributions in statistics (Vol. 2). Wiley.

Liu, S., Wang, Y., Ma, J., & Xu, C. (2023). Identification of road traffic accident hotspots
using machine learning: A case study of Beijing, China. Accident Analysis & Prevention, 180, 107223.

McCullagh, P., & Nelder, J. A. (1989). Generalized linear models (Vol. 37). Chapman and Hall.

National Bureau of Statistics, Nigeria. (2023). Road Transport Statistics. https://www.nigerianstat.gov.ng/

National Highway Traffic Safety Administration. (2022, October). Traffic safety facts: 2020 data. https://www.nhtsa.gov/

Oyedepo, O. J., & Makinde, O. J. (2010). Road traffic accidents in Nigeria: A case study of Ibadan metropolis. Journal of Applied Sciences in Environmental Sanitation, 5(2), 229-234.

Winkelmann, R. (2008). Econometric analysis of count data (5th ed.). Springer.

World Health Organization. (2020). Road traffic injuries. https://iris.who.int/bitstream/handle/10665/44122/9789241563840_eng.pdf