

Test of Significancy and Error Associated with Road Traffic Crashes Data

Odukoya Elijah Ayooluwa, Ayeni Taiwo Michael*, Ilesanmi Anthony Opeyemi, Ogunwale Olukunle Daniel, Aladejana Ayosunkannmi Emmanuel

Ekiti State University, Nigeria

DOI: <https://doi.org/10.51244/IJRSI.2025.12040135>

Received: 08 April 2025; Accepted: 19 April 2025; Published: 21 May 2025

ABSTRACT

This study compared the performance of Poisson regression, negative binomial regression and generalized negative binomial regression models using road traffic crashes data collected from Ekiti state command of Federal Road Safety Commission (FRSC). The method of maximum likelihood estimation was employed to determine the parameter estimate for each models, the significance of each variable consider were examined. The criteria used in selecting the best model fitted are Akaike information criterion (AIC), Bayesian information criterion and Deviance. R software package was used in analysing the data. Firstly multicollinearity was tested to determine if there is high correlation between the predictor variables. Data were analysed and the result from the three model were compared using Akaike information criterion (AIC) and Bayesian information criterion (BIC) and the deviance. The result from the three models were compared using AIC, BIC and Deviance, AIC values for the models are (414.79, 476.8 and 587.312), BIC values (490.8873, 495.59 and 589.321) and Deviance value (61.93, 66.927 and 69.927) respectively. Generalized negative binomial regression (GNBR) had the least value of AIC and BIC, thus consider as the best model for analyzing road traffic crashes data in Ekiti state, Nigeria. Having shown a smaller AIC, BIC and deviance values, the generalized negative binomial was consider a better fit model when analysing road traffic crashes in Ekiti state, Nigeria.

Keywords: Generalized Negative Binomial regression, Negative Binomial Regression Poisson regression, Multicollinearity, Akaike, Bayesian information criterion.

INTRODUCTION

Road traffic crashes require model that will provide good fit for data that will be collected for productive performance. Poisson regression negative binomial regression and generalized negative binomial regression models are techniques used to model dependent variable that describe count data (Cameron and Trivedi 1998).it is often applied to study the occurrence of small number of count as a function of a set predictor variable in experiment and observation study in many discipline, including Economics, Demography psychology, biology and medicine (gardnener 1995) .Poisson regression is a standard model for fitting count data when the number of occurrence of a phenomenon occurred at a constant rate with respect to time but Famoye (2004) noted that Poisson regression is not appropriate when data exhibit over dispersion in which negative binomial take into count the extra-variability(over-dispersion) observed in actual data. Negative binomial loosens the restrictive condition that the variance is equal to the mean, negative binomial regression is used to address the issue of over-dispersion by including the dispersion parameter to accommodate the unobserved heterogeneity in the count data and it that dispersion is large due to the nature of the crashes data which is subjected to Bernoulli trials. Negative binomial is better for over-dispersion count data that are not necessarily heavy-tailed. Poisson gamma can be considered as generalization of Poisson regression as the name implies Poisson gamma is a mixture of two distribution and is first derived by Greenwood and Yule (1920). It became very popular because the conjugate distribution (same family) has a closed form that lead to negative binomial and according to Cook (2009) the name of the distribution comes from applying the binomial theorem with a negative exponent. Generalized negative binomial regression, handle data with both over-dispersion and high frequency of zero. The mean and variance are approximately equal and so

generalized negative binomial regression resemble the Poisson regression, which shows an advantage over the Poisson regression. Generalized negative binomial converges to a Poisson type regression in which the variance may be more or less than the mean, depending upon the value of the parameter. The parameter such that both the variance and mean are positively correlated to the value of the parameter.

LECTURE REVIEW

Statistical models are used to examine the relationships between accidents and features of accidents as well as accident sites. However, many past studies illuminating the numerous problems with linear regression models (Joshua and Garber, 1990 and Miaou and Lum, 1993) haven led to the adoption of more appropriate regression models such as Generalized negative binomial converges to a Poisson type distribution in which the variance may be more or less than the mean, depending upon the value of the parameter. Generalized negative binomial, handle data with both over-dispersion and high frequency of zero, Poisson regression which is used to model data that are Poisson distributed, and negative binomial (NB) model which is used to model data that have gamma distributed Poisson means across crash sites allowing for additional dispersion (variance) of the crash data. Although the Poisson and NB regression models possess desirable distributional properties to describe motor vehicle accidents, these models are not without limitations. One problem that often arises with crash data is the problem of excess 'zeroes, which often leads to dispersion above that described by even the negative binomial model. Excess do not mean too many in the absolute sense, it is a relative comparison that merely suggests that the Poisson and or negative binomial distributions predict fewer zeroes than present in the data. As discussed in Lord et al. (2004), the observance of a preponderance of zero crashes results from low exposure (i.e. train frequency and traffic volumes), high heterogeneity in crashes, observation periods that are relatively small, and or under-reporting of crashes, and not necessarily a dual state 'process which underlies the zero-inflated 'model. Thus, the motivation to compare the three existing models and find better fitting models which from a statistical stand point is justified; unfortunately, however, the zero-inflated model comes also with excess theoretical baggage.

Lacks theoretical appeal (see Lord et al., 2004). Another problem not often observed with crash data is under dispersion, where the variance of the data is less than the expected mean under an assumed probability model (e.g. the generalized negative binomial).

Opong and Asiedu (2014) in their study employed the negative binomial regression to fit a model to the secondary data to ascertain the significance or otherwise of vehicular type to road accidents' fatality in Ghana. The negative binomial regression model was fitted into the data using the Akaike information criterion (AIC) for the best model selection. They observed from their results that in general the number of people killed in road accidents gradually increases with time.

Opong *et al* (2012) compared poisson regression and negative binomial regression by using Poisson regression to fit a model to the secondary data which was obtained from building and road research institute of the council for scientific and industrial research on the number of people killed by road accident in Ghana from 2001 to 2010, given the type of vehicle which was involve in the accident, the age of those who were killed, day of the accident and time in the year. The result of the Poisson analysis showed that there was over dispersion in the data. Negative binomial regression analysis was therefore used to validate the Poisson regression model. it was shown that the negative binomial regression model was the best fit for the data.

Omari-Sasu (2016) several models were compared to determine an appropriate count regression model that adequately fits road accidents in Ghana and determine the key predictors using the appropriate model with respect to the expected number of persons killed in road accidents. Several models were compared to fit count data that encounter the field of transportation. These models include Poisson, Negative Binomial (NB) and Conway-Maxwell-Poisson (CMP) models. In order to compare the performance of these models, the various model selection methods such as Deviance goodness of fit test, Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) were employed. Because the values of Deviance goodness of fit test, AIC and BIC for the NB model was the smallest as compared to that of the Poisson and CMP models, it appeared that, the NB model performed best than the Poisson and CMP models. Base on the appropriate model selected (NB model), the key predictors that contributed significantly and also had a high effect on the

expected or mean number of persons killed in road accidents within a particular period were Head-on collision as Collision type, Improper overtaking and Loss of control as Driver errors, Bus/Minibus as Type of vehicle, Fog/Midst as Weather condition and Night with street lights off as Light condition.

DATA AND METHODOLOGY

In this section, the data used in this study and the methodology are described. Data of road traffic crashes was collected from Federal Road Safety Command from year 2010 to 2018, data on number of people involved, number of people killed, season of the year, total fatal, total minor number of crashes. Data was analyzed by R statistical software package and to access the performance of these models the measures of goodness of fit such as Akaike information criterion and Bayesian information criterion and deviance value was employed. The model with the least Akaike information criterion (AIC), Bayesian information criterion (BIC) value and Deviance value will be considered to be the best model and also the regression estimates for the three models was generated., and muticollinarity test was performed in other to know maybe there are high correlation between two or more predictor variable, which skewing the result in regression model.

Modeling of Crash Data

Poisson and Negative Binomial (NB) regression and generalized negative binomial regression models was compared The negative binomial (NB) regression model is the member of the exponential family of discrete probability distributions. The nature of the distribution is itself well understood,

but its contribution to regression modeling, in particular as a generalized linear model (GLM), has not been appreciated until recently. The mathematical properties of the negative binomial are derived and GLM algorithms are developed for both the canonical and log form. The log forms of both may be effectively used to model types Poisson-over-dispersed count data (Hilbe, 1993). It is not recommended that negative binomial models be applied to small samples. Poisson regression is, also a member of the class of models known as generalized linear models (GLM), is the standard method used to analyze count data. However, many real data situations violate the assumptions upon which the Poisson model is based. For instance, the Poisson model assumes that the mean and variance of the response are identical. This means that events occur within a period of observation at a constant rate, an event is equally likely at any point within the period. When there is heterogeneity in the data, it is likely that the Poisson model is over-dispersed. Such over-dispersion is indicated if the variance of the response is greater than its mean. One may also check for model over-dispersion by submitting the data to a Poisson model. The model is Poisson-over-dispersed if the dispersion value is greater than unity. Negative binomial regression can rather effectively be used to model count data in which the response variance is greater than that of the mean (Hilbe, 1993). A Generalized negative binomial on the other hand reduces to binomial or negative binomial distribution because the mean and variance tend to increase or decrease with an increase or decrease in its value but the variance increase or decrease faster than the mean. The mean and variance are approximately equal and so generalized negative binomial resemble the Poisson distribution, which shows an advantage over the Poisson regression. Generalized negative binomial converges to a Poisson type distribution in which the variance may be more or less than the mean, depending upon the value of the parameter. Generalized negative binomial, handle data with both over-dispersion and high frequency of zero, several models have been developed to explain traffic activities, as of today, the approach is to model how various safety aspects affect the overall road fatality rate (Lindberg *et al.*, 2012)

Model Result

TABLE A: COLLINEARITY STATISTICS

Model	Tolerance	VIF
Season (Month of the Year)	1.0125	0.9876
Number of causes	4.8826	0.2048
Vehicles Involve	4.0477	0.2471

No Injured	1.8804	0.5318
No killed	1.233	0.8108

The Table above shows that all the variables have VIF values < 10, using rule of thumb since they are all lesser than 10, there is no multicollinearity. Thus, all the variable can be Included in the subsequent analyses and modelling with the Poisson regression, Negative Binomial regression and Generalized Negative Binomial.

TABLE B: POISSON REGRESSION

Parameters	Estimate	Standard Error	Z value	Pr(> z)
Intercept	1.016912	0.165990	6.126	8.99e-10
Season (Month of the Year)	-0.008622	0.010712	-0.805	0.42085
Number of causes	0.017069	0.011436	1.493	0.00312
Vehicles Involved	0.045618	0.005069	8.999	< 2e-16
Number Injured	0.011159	0.003550	3.143	0.00167
Number killed	-0.003965	0.014258	-0.278	0.78096

The Table above shows the result of the Poisson regression using the p value in the last column, number of causes vehicles involve, number injured have significant effect on road traffic crashes. While others as no significant effect

$$\eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} \dots \beta_k X_{ik}$$

The regression model, which establishes the relationship between road traffic accidents and month of the year is thus given as $Y = 1.016912 + (-0.008622X_1)$, where Y is the total road traffic accidents and X1 being the season of the year, From the regression model obtained, the value of 1.016912 is interpreted to be the total number of road traffic accidents when the season of the year is set to zero and all other factors are held constant, while the coefficient of X1 of -0.008622 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the season of the month. Its negative sign is an indication of the fact that there is a negative association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant term, as well as that of the predictor variable (X1) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and number of causes is thus given as $Y = 1.016912 + (-0.008622X_1) + 0.017069X_2$, where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes, From the regression model obtained, the value of 1.016912 is interpreted to be the total number of road traffic accidents when the season number of causes is set to zero and all other factors are held constant, whilst the coefficient of X2 of 0.017069 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the number of causes. Its positive sign is an indication of the fact that there is a positive association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X2) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and number of causes is thus given as $Y = 1.016912 + (-0.008622X_1) + 0.017069X_2 + 0.045618X_3$ where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes, X3 vehicles involved From the regression

model obtained, the value of 1.016912 is interpreted to be the total number of road traffic accidents when the vehicles involved is set to zero and all other factors are held constant, whilst the coefficient of X3 of 0.045618 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the vehicles involved. Its positive sign is an indication of the fact that there is a positive association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant term, as well as that of the predictor variable (X2) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and number of causes is thus given as $Y = 1.016912 + (-0.008622X_1) + 0.017069X_2 + 0.045618X_3 + 0.011159X_4$ where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes, X3 vehicles involved X4 number injured From the Poisson regression model obtained, the value of 1.016912 is interpreted to be the total number of road traffic accidents when the number injured is set to zero and all other factors are held constant, whilst the coefficient of X4 of 0.011159 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the number killed. Its positive sign is an indication of the fact that there is a positive association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X4) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and number killed is thus given as $Y = 1.016912 + (-0.008622X_1) + 0.017069X_2 + 0.045618X_3 + 0.011159X_4 + (-0.003965X_5)$ where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes, X3 vehicles involved X4 number injured X5 number killed From the Poisson regression model obtained, the value of 1.016912 is interpreted to be the total number of road traffic accidents when the number injured is set to zero and all other factors are held constant, whilst the coefficient of X5 of -0.003965 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the number killed. Its negative sign is an indication of the fact that there is a negative association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X5) are statistically significant since it is less than the chosen alpha level of 0.05.

TABLE C: NEGATIVE BINOMIAL REGRESSION

Parameters	Estimate	Standard Error	Z value	Pr(> z)
Intercept	1.016903	0.165995	6.126	9.01e-10
Season (Month of the Year)	-0.008622	0.010712	-0.805	0.42087
Number of causes	0.017069	0.011437	1.492	< 2e-16
Vehicles Involve	0.045618	0.005069	8.999	< 2e-16
No Injured	0.011160	0.003551	3.143	0.00167
No killed	-0.003964	0.014258	-0.278	0.78100

Table C shows the result of the negative binomial regression using the p value in the last column number of causes, vehicle involve , number injured have significant effect on road traffic crashes while others have no significant effect.

The regression model, which establishes the relationship between road traffic accidents and season of the year is thus given as $Y = 1.016903 + (-0.008622X_1)$, where Y is the total road traffic accidents and X1 being the moth of the year, From the regression model obtained, the value of 1.016903 is interpreted to be the total number of road traffic accidents when the season of the year is set to zero and all other factors are held

constant, whilst the coefficient of X1 of -0.008622 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the month of the year. Its negative sign is an indication of the fact that there is a negative association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X1) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and number of causes is thus given as $Y = 1.016903 + (-0.008622X_1) + 0.017069X_2$, where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes, From the negative binomial regression model obtained, the value of 1.016903 is interpreted to be the total number of road traffic accidents when the number of causes is set to zero and all other factors are held constant, whilst the coefficient of X2 of 0.017069 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the number of causes. Its positive sign is an indication of the fact that there is a positive association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X2) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and vehicles involved is thus given as $Y = 1.016903 + (-0.008622X_1) + 0.017069X_2 + 0.045618X_3$ where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes, X3 vehicles involved From the regression model obtained, the value of 1.016903 is interpreted to be the total number of road traffic accidents when the vehicles involved is set to zero and all other factors are held constant, whilst the coefficient of X3 of 0.045618 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the vehicles involved. Its positive sign is an indication of the fact that there is a positive association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X2) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and number injured is thus given as $Y = 1.016903 + (-0.008622X_1) + 0.017069X_2 + 0.045618X_3 + 0.011160X_4$ where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes, X3 vehicles involved X4 number injured From the negative binomial regression model obtained, the value of 1.016903 is interpreted to be the total number of road traffic accidents when the number injured is set to zero and all other factors are held constant, whilst the coefficient of X4 of 0.011160 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the number killed. Its positive sign is an indication of the fact that there is a positive association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X4) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and number killed is thus given as $Y = 1.016903 + (-0.008622X_1) + 0.017069X_2 + 0.045618X_3 + 0.011159X_4 + (-0.003964X_5)$ where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes, X3 vehicles involved X4 number injured X5 number killed From the negative binomial regression model obtained, the value of 1.016903 is interpreted to be the total number of road traffic accidents when the number injured is set to zero and all other factors are held constant, whilst the coefficient of X5 of -0.003965 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the number killed. Its negative sign is an indication of the fact that there is a negative association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X5) are statistically significant since it is less than the chosen alpha level of 0.05

TABLE D: GENERALIZED NEGATIVE BINOMIAL

Parameters	Estimate	Standard Error	Z value	Pr(> z)
Intercept	0.4937806	0.0132060	37.391	< 2e-16
Season (Month of the Year)	0.0044566	0.0013295	3.352	0.00113
Number of causes	0.0188368	0.0011600	16.238	< 2e-16
Vehicles Involve	0.0332233	0.0005804	57.237	< 2e-16
No Injured	0.0029701	0.0004539	6.543	2.45e-09
No killed	-0.0185590	0.0011766	-15.773	< 2e-16

Table D shows the result of the generalized negative binomial regression using the p value in the last column season of the year, number of causes, vehicle involve, number injured and no killed have significant effect on road traffic crashes.

The regression model, which establishes the relationship between road traffic accidents and season of the year is thus given as $Y = 0.4937806 + 0.0044566X_1$, where Y is the total road traffic accidents and X1 being the month of the year. From the generalized negative binomial regression model obtained, the value of 0.004566 is interpreted to be the total number of road traffic accidents when the season of the year is set to zero and all other factors are held constant, whilst the coefficient of X1 of 0.0044566 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the month of the year. Its positive sign is an indication of the fact that there is a positive association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X1) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and number of causes is thus given as $Y = 0.4937806 + 0.0044566X_1 + 0.0188368X_2$, where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes. From the generalized negative binomial regression model obtained, the value of 0.4937806 is interpreted to be the total number of road traffic accidents when the number of causes is set to zero and all other factors are held constant, whilst the coefficient of X2 of 0.0188368 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the number of causes. Its positive sign is an indication of the fact that there is a positive association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X2) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and vehicles involved is thus given as $Y = 0.4937806 + 0.0044566X_1 + 0.0188368X_2 + 0.0332233X_3$ where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes, X3 vehicles involved. From the generalized negative binomial regression model obtained, the value of 0.4937806 is interpreted to be the total number of road traffic accidents when the vehicles involved is set to zero and all other factors are held constant, whilst the coefficient of X3 of 0.0332233 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the vehicles involved. Its positive sign is an indication of the fact that there is a positive association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X2) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and number injured is thus given as $Y = 0.4937806 + 0.0044566X_1 + 0.0188368X_2 + 0.00332233X_3 + 0.0029701X_4$ where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes, X3 vehicles involved X4 number injured From the generalized negative binomial regression model obtained, the value of 0.4937806 is interpreted to be the total number of road traffic accidents when the number injured is set to zero and all other factors are held constant, whilst the coefficient of X4 of 0.0029701 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the number killed. Its positive sign is an indication of the fact that there is a positive association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X4) are statistically significant since it is less than the chosen alpha level of 0.05.

The regression model, which establishes the relationship between road traffic accident and number killed is thus given as $Y = 0.4937806 + 0.0044566X_1 + 0.0188368X_2 + 0.0332233X_3 + 0.0029701X_4 + (-0.0185590X_5)$ where Y is the total road traffic accidents and X1 being the season of the year X2 number of causes, X3 vehicles involved X4 number injured X5 number killed From the generalized negative binomial regression model obtained, the value of 0.4937806 is interpreted to be the total number of road traffic accidents when the number injured is set to zero and all other factors are held constant, whilst the coefficient of X5 of -0.0185590 is the rate or magnitude of change in the road traffic accident as a result of a unit change in the number killed. Its negative sign is an indication of the fact that there is a negative association between road traffic accidents. Again, the errors associated with these coefficients are minimal as displayed by the standard errors of the coefficients. The p-values indicate that the constant terms, as well as that of the predictor variable (X5) are statistically significant since it is less than the chosen alpha level of 0.05.

TABLE E: AIC, BIC AND DEVIANCE VALUES FOR THE THREE MODELS

Model	AIC	BIC	Deviance
Generalized negative binomial regression	414.79	460.8873	61.93
Negative binomial	476.8	495.57	66.927
Poisson regression	587.3196	589.3208	69.927

Comparison using AIC and BIC values in table B, D and F, the estimated AIC and BIC for the Poisson regression is 587.3196 and 589.3208 respectively, whereas it is 476.8 and 495.57 for Negative binomial and for Generalize Negative binomial is 414.79 and 460.8873 respectively. The smallest value of AIC and BIC is the generalized negative binomial regression and therefore it is the optimal model.

FINDINGS AND CONCLUSION

Road traffic crashes are count (discrete) in nature. In modeling discrete data for characteristics and prediction of an event when dependent variables are non-negative with integer's values. It is appropriate to use Poisson regression. Poisson regression is a standard model for fitting count data when the number of occurrences of a phenomenon occur at a constant rate in respect to time and an occurrence of any future occurrence.

However, the condition that mean and variance of Poisson regression are equal to each other poses a great constraint. This bring about the use of the Generalized negative binomial regression (GNB) and Negative binomial regression (NBR) models, to overcome the earlier restrictive condition that the variance is equal to the mean. This study compared the performance of Poisson regression, negative binomial regression and generalized negative binomial regression models using road traffic crashes data collected from Ekiti state command of Federal Road Safety Commission (FRSC).

The method of maximum likelihood estimation was employed to determine the parameter estimate for each model; the significance of each variable consider were examined. The criteria used in selecting the best model fitted are Akaike information criterion (AIC), Bayesian information criterion and Deviance. R software

package was used in analysing the data. Firstly, multicollinearity was tested to determine if there is high correlation between the predictor variables. Data were analysed and the result from the three models were compared using Akaike information criterion (AIC) and Bayesian information criterion (BIC) and the deviance. Generalized negative binomial showing an AIC value of 414.79, BIC value of 490.8873 and Deviance value 61.99 and negative binomial regression showing AIC value of 476.8, BIC value of 495.57 and Deviance value 66.927 with Poisson regression showing AIC value of 587.3196, BIC value of 589.3208 and Deviance value 69.927. Having shown a smaller AIC and BIC value, the Generalized negative binomial was considered a better model when analysing road traffic crashes in Ekiti state, Nigeria.

Having compared the three models on road traffic crashes in Ekiti state Nigeria and the result from the three model were compared using AIC and BIC and with Generalized negative binomial having the smaller AIC and BIC, Generalized negative binomial was considered as a best model when analysing road traffic crashes in Ekiti State Nigeria,

REFERENCE

1. Akaike H., (1973). Information theory and the extension of likelihood principle. Proceeding of the international symposium on information theory. Akademiakaidoo, Budapest Hungary 2266-282
2. Cook, J.D. (2009). Notes on the Negative Binomial Distribution. http://www.johndcook.com/negative_binomial.pdf. (Accessed on Aug. 15th, 2010)
3. Famoye F, John T. W. and Karan P. S. (2004), "On the generalized Poisson regression with an application to accident data: Journal of Data Science 2(2004), 287-295
4. Gardner, W., Mulvey E. P., Shaw E. C. (1995). Regression analyses of counts and rates: Poisson, over dispersed Poisson, and negative binomial models. Psychological Bulletin 118: 392-404.
5. Greenwood, M. and Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attack of disease or repeated accidents. Journal of the Royal Statistical Society 83(1), 255-279.
6. Hilbe, J.M. (1993). Negative Binomial Regression as a Generalized Linear Model.
7. Technical Report COS 93/94-5-26. Department of Sociology and Graduate College
8. Joshua, S.C. and Garber, N.J. (1990). Estimating truck accident rate and involvements
9. Using linear and Poisson regression models. Transport. Planning Technol. 15 (1990)(1), pp. 41-58.
10. Lindberg, J., Darin, L. and Berg, Y. 2012. Review of road goals and indicators for road safety between 2010 and 2020. Swedish Transport Administration
11. Miaou S. P., Lum H. (1993a). Modeling vehicle accidents and highway geometric design relationships Accident Analysis and Prevention. 25:689-1993 [http://dx.doi.org/10.1016/0001-4575\(93\)90034-t](http://dx.doi.org/10.1016/0001-4575(93)90034-t)
12. Omari-Sasu, A.Y., Adjei, M.I., Boadi, R.K., (2016) Statistical Models for Count Data with Applications to Road Accidents in Ghana International. Journal of Statistics and Applications 2016, 6(3): 123-137 DOI: 10.5923/j.statistics.20160603.0
13. Oppong R. A. (2014). Analysis of age as a risk factor of road accidents fatality in Ghana with negative binomial model. International Journal of Modern Sciences and Engineering Technology 2(12); 2634-2649 ISSN (e): 2321-7545 <http://ijsae.in>
14. Oppong R. A., Asiedu-Addo S. K. (2014). Analysis of vehicular type as a risk factor of road accidents fatality in Ghana. International Journal of Modern Science and Engineering Technology 1(5): 106-114
15. Oppong, R.A., Assuah, C.K., Asiedu-Addo S.K. (2015). Comparative Assessment of Poisson and Negative Binomial Regressions as Best Models for Road Count Data; International Journal of Scientific Research and Engineering Studies (IJSRES), 2(11);28-32
16. Winkelmann, R. and Zimmermann, K. (1995). Recent developments in count data Modeling: theory and applications. J. Econ. Surveys 9, pp. 1-24