Statistical Model Of Road Traffic Crashes Data In Anambra State, Nigeria: A Poisson Regression Approach

Nwankwo Chike H., Nwaigwe Godwin I

Abstract: Road traffic crashes are count (discrete) in nature. When modeling discrete data for characteristics and prediction of events, it is appropriate using the Poisson Regression Model. However, the condition that the mean and variance of the Poisson are equal, poses a great constraint, hence necessitating the use of the Generalized Poisson Regression (GPR) and the Negative Binomial Regression (NBR) models, which do not require these constraints that the mean and the variance be equal, as proxies. Data on Road traffic crashes from the Anambra State Command of the Federal Road Safety Commission (FRSC), Nigeria were analyzed using these three methods, the results from the two proxies are compared using the Akaike Information Criterion (AIC) with GPR showing an AIC value of 3508.595 and the NBR showing an AIC value of 2742. Having shown a smaller AIC value, the NBR was considered a better model when analyzing road traffic crashes in Anambra State, Nigeria.

Keywords: Over-dispersion, Road Traffic, Crashes, Discrete, Akaike Information Criterion.

1.0 INTRODUCTION

Road traffic crashes are of grave concern as many of them end up taking life. The morbidity and mortality burden is rising in developing countries due to a combination of factors, including rapid motorization, poor road and traffic infrastructure, as well as the behavior of road users (Nantulya and Reich, 2002). This contrasts with technologically advanced countries where the indices are reducing (Oskam et al., 1994; O'Neill and Mohan, 2002). Anambra state, a slight heavily motorized state in Nigeria with poor road conditions and transport systems has a high rate of road traffic crashes and the tendency is yet unabated. The recognition of road traffic crashes as a crisis in Nigeria inspired the establishment of the Federal Road Safety Commission (FRSC). The FRSC was established by the government of the Federal Republic of Nigeria vide Decree 45 of 1988 as amended by Decree 35 of 1992, with effect from 18th February, 1998. The commission was charged with responsibilities of, among others, policymaking, organization and administration of road safety in Nigeria. Famoye, et al (2004), noted that the Poisson regression model is not appropriate when a data set exhibit over-dispersion, a condition where the variance is more than the mean. This overdispersion is almost always the case when the Poisson regression models are adopted for modeling such data.

2.0 METHODS

Road traffic crashes, number of people involved, causes and number of death recorded in road traffic crashes are all discrete or count data. The data are collected over a fixed continuous space which is Anambra state, Nigeria and over a fixed time. Data of this nature are considered as Poisson distributed (Horim and Levy; 1981). This makes it appropriate to use the Poisson distribution in modeling the data on road traffic crashes in Anambra state, Nigeria

2.0.1 POISSON REGRESSION MODEL

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

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

And two functions

- A Link function that describes how the mean, $E(Y_i) = \mu_i$, depends on the linear predictor $g(\mu_i) = \eta_i$ (2)
- A variance function that describes how the variance, $\mathrm{var}(Y_i)$ depends on the mean $\mathrm{var}(Y_i) = \phi var(\mu)$ (3)

Where the dispersion parameter ϕ is a constant.

Suppose $Y_i \sim \text{Poisson}\ (\lambda_i)$

Then,

$$E(Y_i) = \lambda_i \qquad \mathrm{var}(Y_i) = \lambda_i \quad (4)$$

Therefore, our variance function is

$$V(\mu_i) = \mu_i \quad (5)$$

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

$$g(\mu_i) = \log_e(\mu_i) \quad (6)$$

The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function. Link function here is the function that links the linear model in a design matrix and the Poisson distribution function. Consider a linear regression model given as

$$y_i = \beta_0 + \beta_i x_i + \varepsilon_i \qquad i = 1, \dots, k \quad (7)$$

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

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

- *Nwankwo Chike H, currently a senior lecturer and a PhD holder in the Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria.*
- *Nwaigwe Godwin I, currently running his masters degree in the Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria.Email of Corresponding Author: brainman4all@yahoo.com*

- *This paper aims at choosing a mathematical model that will help understand and man*age road traffic *crashes in Anambra state, Nigeria.*

Where
X is an $n \times (K+1)$ vector of independent variables or predictors, and a column of 1's
β is a $(k+1)$ by 1 vector of unknown parameters and
ε is an $n \times 1$ vector of random error terms with zero mean
Hence,

$$\mathrm{E}(Y/X) = X\beta \quad (9)$$

Recall that, for Generalized Linear Models, we use the link function to transform Y:
That is,

$$G(\mathrm{Y}) = \log_e(\mathrm{Y}) \quad (10)$$

Therefore, this can be written more compactly as

$$\log_e \mathrm{E}(Y/X) = x\beta \quad (11)$$

Thus, given a Poisson regression model with parameter β and an input vector x, the predicted mean of the associated Poisson distribution is given by

$$\mathrm{E}(y/x) = \mathrm{e}^{x\beta} \quad (12)$$

If y_i, are independent observations with corresponding values x_i of the predicted variables, then β can be estimated by maximum likelihood. The maximum-likelihood estimates lack a closed-form expression and must be found by numerical methods. The probability surface for maximum-likelihood Poisson regression is always convex, making Newton-Raphson or other gradient-based methods appropriate estimation techniques. Therefore, let y_i be the random variable, that takes non-negative values, $i = 1, 2, \dots, n$, *Where n is the number of observations.* Since y_i follows a Poisson distribution, the probability mass function (pmf) is

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

With mean and variance as

$$\mathrm{E}(Y_i) = Var(Y_i) = \lambda_i \quad (14)$$

where the conditional mean or predicted mean of the poisson distribution as given in (12) above specified by

$\mathrm{E}(y/x) = \mathrm{e}^{x\beta} = \lambda_i$ *(Same as the mean of the Poisson)*
x *is the value of the explanatory variable and*
$\beta = (\beta_1, \beta_2, \dots, \beta_k)$ *are the unknown* $k-$*dimensional vector of regression parameters*

The mean of the predicted Poisson distribution is given by $\mathrm{E}(y/x)$ and variance of y_i as $\mathrm{var}(y/x)$. The parameters β can be estimated by maximum likelihood estimation method

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

The log-likelihood function is given by

$$\ln \ell(\beta) = \sum_{i=1}^{n} [-\lambda_i + y_i \ln \lambda_i - \ln y_i!] \quad (16)$$

Substituting $\lambda_i = \mathrm{e}^{x\beta}$ in the equation above, we have

$$= \sum_{i=1}^{n} \left[y_i(x\beta) - \mathrm{e}^{x\beta} - \ln y_i! \right] \quad (17)$$

Differentiating equation (17) with respect to β, and equating to zero, we have

$$\frac{\partial \ln \ell(\beta)}{\partial \beta} = \sum_{i=1}^{n} \left(y_i - exp(x\beta) \right) x = 0 \qquad i = 1, 2, \dots, k \quad (18)$$

Which yields k non-linear equations and can be solved using the Newton-Raphson method.

2.0.2 NEGATIVE BINOMIAL REGRESSION (NBR)

One important application of the negative binomial distribution is that it is a mixture of a family of Poisson distributions with gamma mixing weights. Thus the negative binomial distribution can be viewed as a generalization of the Poisson distribution. The negative binomial can be viewed as a Poisson distribution where the Poisson parameter is itself a random variable, distributed according to a Gamma distribution. Thus the negative binomial distribution is known as a Poisson-Gamma mixture. As the most common alternative to Poisson regression, the negative binomial regression addresses the issue of over-dispersion by including a dispersion parameter to accommodate the unobserved heterogeneity in the count data. The negative binomial regression (Poisson-Gamma) can also be considered a generalization of Poisson regression. As its name implied, the negative binomial (Poisson-gamma) is a mixture of two distributions and was first derived by Greenwood and Yule (1920). It became very popular because the conjugate distribution (same family of functions) has a closed form and leads to the negative binomial distribution. As discussed by Cook (2009), "the name of this distribution comes from applying the binomial theorem with a negative exponent". This mixture distribution was developed to account for over-dispersion that is commonly observed in discrete or count data (Lord et al (2005)).

The model

Suppose that we have a series of random counts that follows the Poisson distribution:

$$f(y_i; \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{y}}{y_i!}, \qquad y \geq 0, \ \lambda \geq 0 \quad (19)$$

Where y_i is the observed number of counts for i=1,2,…,n ,
λ_i is the mean of the Poisson distribution. If the mean is assumed to have a random intercept term and this term enters the conditional mean function in a multiplicative manner, we get the following relationship (Cameron and Trivedi, (1998))

$$\lambda_i = \exp\left(\beta_o + \sum_{j=1}^{k} x_{ij} \beta_j + \varepsilon_i\right) \quad (20)$$

$$\lambda_i = \mathrm{e}^{\sum_{j=1}^{k} x_{ij}\beta_j} \mathrm{e}^{(\beta_o + \varepsilon_i)}$$

$$\lambda_i = \mathrm{e}^{(\beta_o + \sum_{j=1}^{k} x_{ij}\beta_j)} \mathrm{e}^{(\varepsilon_i)}$$

$$\lambda_i = \mu_i \nu_i \quad (21)$$

Where $\mathrm{e}^{(\varepsilon_i)} \sim gamma\ (\alpha^{-1}; \alpha^{-1})$; $\mathrm{e}^{(\beta_o + \varepsilon_i)}$ is defined as a random intercept;
$\mu_i = \exp\left(\beta_o + \sum_{j=1}^{k} x_{ij} \beta_j\right)$ is the log-link between the Poisson mean and the covariates or independent variables $x's$; and $\beta's$ are the regression coefficients. The marginal distribution of y_i can be obtained by integrating the error term, ν_i,

$$\therefore\ f(y_i; \mu_i) = \int_0^\infty g(y_i; \mu_i, \nu_i) h(\nu_i)\, d\nu_i \quad (22)$$

$$f(y_i; \mu_i) = \mathrm{E}_\nu[g(y_i; \mu_i, \nu_i)]$$

Where $h(\nu_i)$ *is a mixing distribution. In the case of the Poisson-Gamma mixture,* $g(y_i; \mu_i, \nu_i)$ *is the Poisson distribution and* $h(\nu_i)$ *is the Gamma distribution. This distribution has a closed form and leads to the NB distribution*
Assume that the variance ν_i follows a two parameter gamma distribution;

$$h(\nu_i; \alpha, \delta) = \frac{\delta^\alpha}{\Gamma(\alpha)} \nu_i^{\alpha-1} \mathrm{e}^{-\nu_i \delta}, \qquad \alpha > 0, \delta > 0, \nu_i > 0 \quad (23)$$

Where $\mathrm{E}(\nu_i) = \alpha/\delta$ and $\mathrm{var}[\nu_i] = \alpha/\delta^2$
Setting □=δ, we have

$$\mathrm{E}(\nu_i) = 1 \quad \text{and} \quad \mathrm{var}[\nu_i] = 1/\alpha \quad (24)$$

We can transform the gamma distribution as a function of the Poisson mean, which gives the following probability density function (PDF; Cameron and Trivedi, (1998)):

$$h(\lambda_i;\alpha,\mu_i)=\frac{\left(\alpha/\mu_i\right)^{\alpha}}{\Gamma(\alpha)}\lambda_i^{\alpha-1}\mathrm{e}^{-\frac{\lambda_i}{\mu_i}\delta} \quad (25)$$

Combining equations (19) and (25) into equation (22) yields the marginal distribution of y_i:

$$f(y_i;\alpha,\mu_i)=\int_0^{\infty}\frac{exp(-\lambda)\lambda_i^{y_i}}{y_i!}\frac{\left(\alpha/\mu_i\right)^{\alpha}}{\Gamma(\alpha)}\lambda_i^{\alpha-1}\mathrm{e}^{-\frac{\lambda_i}{\mu_i}\delta}d\lambda_i \quad (26)$$

Using the properties of the gamma function and Setting δ=□, it can be shown that equation (26) can be defined as;

$$f(y_i;\alpha,\mu_i)=\frac{\left(\alpha/\mu_i\right)^{\alpha}}{\Gamma(\alpha)\Gamma(y_i+1)}\int_0^{\infty}exp\left(-\lambda_i\left(1+\alpha/\mu_i\right)\right)\lambda_i^{y_i+\alpha-1}d\lambda_i$$

(27)

$$f(y_i;\alpha,\mu_i)=\frac{\left(\alpha/\mu_i\right)^{\alpha}\left(1+\alpha/\mu_i\right)^{-(\alpha+y_i)}\Gamma(\alpha+y_i)}{\Gamma(\alpha)\Gamma(y_i+1)}$$

$$=\frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)\Gamma(y_i+1)}\left[\frac{\alpha}{\mu_i+\alpha}\right]^{\alpha}\left[\frac{\mu_i}{\mu_i+\alpha}\right]^{y_i} \quad (28)$$

Therefore, the pdf of the negative binomial regression is

$$f(y_i;\alpha,\mu_i)=\frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)\Gamma(y_i+1)}\left[\frac{\alpha}{\mu_i+\alpha}\right]^{\alpha}\left[\frac{\mu_i}{\mu_i+\alpha}\right]^{y_i} \quad (29)$$

The mean and variance are the following (Lord and Park (2014))

$$E(y_i;\alpha,\mu_i)=\mu_i \quad (30)$$

$$Var(y_i;\alpha,\mu_i)=\mu_i+\frac{\mu_i^2}{\alpha} \quad (31)$$

The next steps consist of defining the log-likelihood function and it can be shown that, (Lord and Park (2013))

$$\ln\left[\frac{\Gamma(y_i+\alpha)}{\Gamma(\alpha)}\right]=\sum_{j=0}^{y-1}\ln(j+\alpha) \quad (32)$$

By substituting equation (32) into (29), the log-likelihood can be computed using the equation

$$\ln L(\alpha,\beta)=\sum_{i=1}^{n}\left\{\left[\sum_{j=0}^{y-i}\ln(j+\alpha)\right]-\ln y_i!-(y_i+\alpha)\ln(1+\alpha^{-1}\mu_i)+y_i\ln\alpha^{-1}+y_i\ln\mu_i\right\}$$

Therefore, the log-likelihood function becomes (Lord and Park (2013));

$$\ln\mathrm{L}(\alpha,\beta)=\sum_{i=1}^{n}\left\{y_i\ln\left[\frac{\alpha\mu_i}{1+\alpha\mu_i}\right]-\alpha^{-1}\ln(1+\alpha\mu_i)+\ln\Gamma(y_i+\alpha^{-1})-\ln\Gamma(y+1)-\ln\Gamma(\alpha^{-1})\right\} \quad (33)$$

2.0.3 THE GENERALIZED POISSON REGRESSION MODEL

The advantage of using the generalized Poisson regression model, is that it can be fitted for both over-dispersion, Var(y_i)>E(y_i), as well as under-dispersion, Var(y_i)<E(y_i) (Wang and Famoye (1997)). Suppose y_i is a count response variable that follows a generalized Poisson distribution, the probability density function of $y_i, i=1,2,\dots,n$ is given as (Famoye (1993), Wang and Famoye (1997));

$$f(y_i)=P(Y_i=y_i)=\left[\frac{\mu_i}{1+\alpha\mu_i}\right]^{y_i}\frac{(1+\alpha y_i)^{y_i-1}}{y_i!}exp\left[-\frac{\mu_i(1+\alpha y_i)}{(1+\alpha\mu_i)}\right],\quad y_i=0,1\dots \quad (34)$$

With mean $\mathrm{E}(y_i)=\mu_i$ *and variance* $Var(y_i)=\mu_i(1+\alpha\mu_i)^2$, Where □ is called the dispersion parameter. The generalized Poisson distribution is a natural extension of the Poisson distribution. If □=0, equation (34) reduces to the Poisson (as in equation 13 above), resulting to Var(y_i) = E(y_i). If □>0, it means Var(y_i)>E(y_i), and the distribution represents count data with over-dispersion. If □<0, it means Var(y_i)<E(y_i), the distribution represents count data with under-dispersion. If it is assumed that the mean or the fitted values is multiplicative, (I.e.) $\mathrm{E}(y_i/x_i)=\mu_i=e_i\exp(x_i\beta)$

Where e_i denotes a measure of exposure, x_i a $pX1$ vector of explanatory variables, and β a $pX1$ vector of regression parameters. The log-likelihood functions of the GPR model is given by (Wang and Famoye (1997);

$$\ell(\beta,\alpha)=\sum_{i=1}^{n}y_i\log\left[\frac{\mu_i}{1+\alpha\mu_i}\right]+(y_i-1)\log(1+\alpha y_i)-\frac{\mu_i(1+\alpha y_i)}{1+\alpha\mu_i}-\log(y_i) \quad (35)$$

Therefore, the maximum likelihood estimates,$(\hat{\alpha},\hat{\beta})$, may be obtained by maximizing $\ell(\beta,\alpha)$ with respect to β and □. The related equations are as follows, Substituting $\mathrm{E}(y_i/x_i)=\mu_i=\exp(x_i\beta)$, *the partial derivative for* β *is,*

$$\frac{\partial\ell(\beta,\alpha)}{\partial\beta}=\sum_{i=1}^{n}\frac{(y_i-\mu_i)\mu_i x_{ij}}{(1+\alpha\mu_i)^2}=0 \qquad j=1,2,\dots,p \quad (36)$$

And

$$\frac{\partial\ell(\beta,\alpha)}{\partial\alpha}=\sum_{i=1}^{n}-\frac{y_i\mu_i}{1+\alpha\mu_i}+\frac{y_i(y_i-1)}{1+\alpha y_i}-\frac{(y_i-\mu_i)\mu_i}{(1+\alpha\mu_i)^2}=0 \quad (37)$$

The parameters □ and β are estimated by the Newton-Raphson method. We can also estimate □ by using method of moments, (i.e.) by equating the Pearson chi-squares statistic with (n-p) degree of freedom, as suggested by Breslow (1995), given as;

$$\sum_{i=1}^{n}\frac{(y_i-\mu_i)^2}{\mu_i(1+\alpha\mu_i)^2}=n-p \quad (38)$$

Where n is the number of values and P is the number of regression parameters.

2.0.4 MULTICOLLINEARITY TEST

Multicollinearity (also collinearity) is a statistical phenomenon in which two or more predictor variables in a multiple regression model are highly correlated, meaning that one can be linearly predicted from the others with a non-trivial degree. In other words, multicollinearity is said to have occurred if two or more independent variables are highly correlated. The multicollinearity test is done as an initial assumption for parameter estimation. One formal way of detecting multicollinearity is by the use of the variance inflation factors (VIF). The VIF is used to test for the presence of multicollinearity, and is given by

$$VIF=\frac{1}{1-R_j^2} \quad (39)$$

Where R_j^2 is the coefficient of determination of a regression of an explanatory variable j on all the other explanators. A VIF value of 10 and above indicates a multicollinearity problem (Wikipedia.org).

2.0.5 AKAIKE INFORMATION CRITERION (AIC)

The Akaike information criterion (AIC) is a measure of the relative quality of a statistical model for a given set of data. That is, given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Hence, AIC provides a means for model selection. Given a set of candidate models for the data, the preferred model is the one with the minimum AIC value (i.e, the smaller the AIC value, the better the model). Hence, AIC does not only reward goodness-of-fit, but also includes a penalty that is an increasing function of the number of estimated parameters (Wikipedia.org). This measure also uses the log-likelihood function, but add a penalizing term associated with number of variables. It is well known that by adding variables, one can improve the fit of models. Thus, the AIC tries to balance the goodness-of-fit versus the inclusion of variables in the model. The AIC is computed as (Wikipedia.org).

$$AIC = 2P - 2\ln L \quad (40)$$

Where P is the number of parameters in the model
L is the maximized value of the likelihood function for the model.

3.0 DATA PRESENTATION

The Causes of accidents as classified by the Federal Road Safety Commission in Nigeria, are listed below;

1. Speed Violation (SPV)
2. Loss of Control (LOC)
3. Dangerous Driving (DGD)
4. Tyre Burst (TBT)
5. Brake Failure (BFL)
6. Wrongful Overtaking (WOT)
7. Route Violation (RTV)
8. Mechanically Deficient Vehicle (MDV)
9. Bad Road (BRD)
10. Road Obstruction Violation (OBS)
11. Dangerous Overtaking (DOT)
12. Overloading (OVL)
13. Sleeping on Steering (SOS)
14. Driving Under Alcohol/Drug Influence (DAD)
15. Use of Phone While Driving (UPWD)
16. Fatigue (FTQ)
17. Poor Weather (PWR)
18. Sign Light Violation (SLV)
19. Others

An accident could be by one or multiple causes as listed above. Where 1 is indicated in the data presentation, it means the cause of the accident is one and could be either of the causes listed above. And where ≥ 2 is indicated in the table, it means that there are 2 or more causes of accidents and so on. The data is presented as in table 1 below:

Table I: *Data on road traffic crashes, seasons of the year and cause of crashes as obtained from the records of the Federal Road Safety Commission, Anambra State Command, Nigeria.*

NUMBER OF CRASHES	PERSONS INVOLVED	SEASON (WEEKS OF THE YEAR)	NUMBER CAUSES
4	6	1	2
7	20	2	4
2	10	3	1
0	0	4	0
0	0	5	0
0	0	6	0
3	20	7	2
0	0	8	0
0	0	9	0
0	0	10	0
0	0	11	0
0	0	12	0
4	3	13	2
4	11	14	2
2	2	15	1
1	1	16	1
0	0	17	0
0	0	18	0
2	6	19	2
6	5	20	3
0	0	21	0
0	0	22	0
0	0	23	0
3	5	24	3
2	6	25	2
0	0	26	0
2	4	27	2
1	3	28	1
1	7	29	1
1	2	30	1
2	4	31	1
1	3	32	1
0	0	33	0
1	3	34	1
2	8	35	1
0	0	36	0
2	10	37	2
1	6	38	1
1	2	39	1
4	6	40	3
7	18	41	5
2	8	42	2
1	33	43	1
2	9	44	2
3	8	45	2
3	2	46	3
1	5	47	1
2	6	48	2
1	1	49	1
4	24	50	4
4	16	51	4
7	8	52	3
1	4	1	1
2	2	2	2
3	35	3	3
4	16	4	4
7	17	5	3
1	7	6	1
1	18	7	1
1	3	8	1
1	2	9	1
3	2	10	2
1	1	11	1
2	1	12	2
2	3	13	2
2	3	14	2
7	26	15	7
5	16	16	5
2	11	17	2

1	8	18	1
1	7	19	1
2	5	20	2
2	24	21	2
1	15	22	1
1	11	23	1
2	2	24	2
7	1	25	3
5	6	26	5
2	7	27	2
2	8	28	2
2	3	29	2
1	12	30	1
4	25	31	4
7	22	32	7
8	18	33	3
1	3	34	1
2	8	35	2
3	16	36	3
4	14	37	4
4	40	38	4
2	20	39	2
1	22	40	1
7	8	41	7
4	13	42	4
3	9	43	3
1	8	44	1
2	10	45	2
2	17	46	2
1	16	47	1
3	6	48	3
3	9	49	3
2	22	50	2
1	18	51	1
2	35	52	2
8	26	1	8
5	3	2	5
1	8	3	1
2	3	4	2
1	7	5	1
2	2	6	2
1	6	7	1
1	5	8	1
0	0	9	0
0	0	10	0
2	10	11	2
4	9	12	4
4	11	13	4
1	5	14	1
2	10	15	2
3	19	16	3
2	14	17	2
6	7	18	6
5	21	19	5
3	5	20	3
4	7	21	4
7	3	22	7
2	7	23	2
1	2	24	1
3	3	25	3
2	5	26	2

2	6	27	2
0	0	28	0
0	0	29	0
3	7	30	3
2	8	31	2
0	0	32	0
3	10	33	3
3	6	34	3
1	2	35	1
1	6	36	1
2	8	37	2
4	33	38	4
1	31	39	1
2	37	40	2
4	21	41	4
4	9	42	4
1	17	43	1
4	30	44	4
1	9	45	1
8	46	46	2
2	10	47	2
6	21	48	2
8	17	49	2
3	23	50	3
2	8	51	2
3	15	52	3
8	7	1	4
4	15	2	4
5	16	3	5
1	15	4	1
4	29	5	4
1	18	6	1
11	32	7	3
4	22	8	2
11	9	9	3
2	45	10	2
3	10	11	3
2	14	12	2
8	21	13	3
7	36	14	7
8	25	15	3
1	14	16	1
2	12	17	2
2	23	18	2
9	24	19	3
1	16	20	1
6	44	21	6
8	18	22	1
5	19	23	5
3	61	24	3
10	12	25	4
6	11	26	6
7	14	27	4
5	16	28	5
8	25	29	8
3	13	30	3
5	22	31	5
2	19	32	2
8	16	33	8
7	53	34	7
9	35	35	3

9	14	36	4
10	32	37	4
6	7	38	6
8	18	39	3
4	42	40	4
1	19	41	1
3	20	42	3
4	14	43	4
6	15	44	6
7	32	45	7
8	8	46	2
8	8	47	3
4	5	48	4
8	8	49	4
8	13	50	4
7	14	51	7
4	16	52	4
5	47	1	5
4	23	2	4
10	19	3	3
6	11	4	2
3	9	5	3
6	23	6	3
1	30	7	1
1	71	8	1
7	65	9	7
7	7	10	7
7	55	11	7
6	8	12	6
2	46	13	2
1	76	14	1
3	3	15	3
3	67	16	3
1	10	17	1
8	18	18	8
3	10	19	3
2	30	20	2
0	0	21	0
10	6	22	3
11	61	23	3
5	47	24	5
1	67	25	1
4	35	26	4
6	81	27	6
4	67	28	4
7	46	29	7
2	12	30	2
5	32	31	5
4	11	32	4
4	41	33	4
2	22	34	2
4	44	35	4
2	16	36	2
0	0	37	0
3	12	38	3
2	29	39	2
5	25	40	5
2	31	41	2
6	15	42	6
1	4	43	1
1	5	44	1

5	84	45	5
6	45	46	6
8	61	47	8
7	27	48	7
0	0	49	0
3	35	50	3
10	20	51	3
6	72	52	6
4	44	1	4
2	35	2	2
2	10	3	2
2	33	4	2
6	85	5	6
2	10	6	2
1	42	7	1
0	0	8	0
3	14	9	3
2	11	10	2
11	95	11	3
10	82	12	4
11	45	13	5
5	61	14	1
10	47	15	4
7	70	16	2
8	32	17	3
8	7	18	2
7	18	19	3
4	42	20	4
5	19	21	5
4	20	22	4
10	9	23	3
6	11	24	2
6	9	25	2
1	23	26	1
1	30	27	1
3	71	28	3
9	53	29	3
4	12	30	4
0	0	31	0
2	7	32	2
4	31	33	4
0	0	34	0
7	33	35	3
2	8	36	2
2	7	37	2
4	50	38	4
2	33	39	2
2	5	40	2
1	2	41	1
7	32	42	2
1	3	43	1
1	6	44	1
1	4	45	1
2	4	46	2
1	1	47	1
0	0	48	0
8	85	49	2
4	44	50	1
1	3	51	1
3	16	52	3
5	58	1	5

0	0	2	0
0	0	3	0
9	66	4	3
0	0	5	0
5	26	6	1
4	52	7	4
9	42	8	6
3	24	9	3
2	30	10	2
16	81	11	9
7	34	12	7
3	5	13	3
3	19	14	3
1	11	15	1
0	0	16	0
1	17	17	1
6	70	18	6
3	15	19	3
2	3	20	2
3	22	21	3
6	73	22	6
1	2	23	1
3	9	24	3
6	23	25	6
4	40	26	4
2	6	27	2
3	38	28	3
3	52	29	3
3	7	30	3
0	0	31	0
5	48	32	3
2	25	33	2
5	13	34	2
2	4	35	2
4	18	36	2
0	0	37	0
0	0	38	0
10	55	39	4
4	30	40	4
3	9	41	3
2	3	42	2
3	27	43	3
0	0	44	0
2	23	45	7

4.0 ANALYSIS AND RESULTS

The data were analyzed using R Software and the results obtained are given below. Before performing the analysis on the three methods used, testing the data for multicollinearity was conducted. The test results are shown in table A below:

***Table A:** Collinearity Statistics*

Model	Tolerance	VIF
Number of Crashes	0.486	2.059
Season (weeks of the year)	0.995	1.005
Number of causes	0.485	2.063

Tabel A shows that all the variables have VIF values <10. Thus all the variables can be included in the subsequent analyses and modelling with the Poisson regression, Generalized Poisson regression, and Negative Binomial Regression. The results from Poisson regression are as follows;

***Table B:** Poisson Regression Model Parameter Estimation*

Parameters	Estimate	Standard Error	z value	Pr(>|z|)
Intercept	2.2432614	0.0329304	68.121	< 2e-16
Number of Crashes	0.1020011	0.0048541	21.013	< 2e-16
Season (weeks of the year)	- 0.0018494	0.0008235	-2.246	0.0247
Number of Causes	0.1030262	0.0074039	13.915	<2e-16

***Table C:** AIC Value of Poisson Regression Model Parameter Estimation*

DEVIANCE	6456.9
AIC	6266

***Table D:** Negative Binomial Regression Model Parameter Estimation*

Parameters	Estimate	Standard Error	z value	Pr(>|z|)
Intercept	2.017437	0.128009	15.760	< 2e-16
Number of Crashes	0.119212	0.025182	4.734	2.20e-06
Season (weeks of the year)	- 0.002007	0.003494	-0.574	0.566
Number of Causes	0.155004	0.037553	4.128	3.67e-05

***Table E:** AIC Value of Negative Binomial Regression Model Parameter Estimation*

DEVIANCE	530.58
AIC	2742

***Table F:** Generalized Poisson Regression Model Parameter Estimation*

Parameters	Estimate	Standard Error	z value	Pr(>|z|)
Intercept	2.1462560	0.0782676	27.4220	< 2e-16
Number of Crashes	0.1005432	0.0178925	5.6193	< 2e-16
Season (weeks of the year)	0.0033059	0.0028278	1.1691	0.00710

Number of Causes	0.1237426	0.0264850	4.6722	0.0841

Table G: *AIC Value of Generalized Poisson Regression Model Parameter Estimation*

DEVIANCE	4.2710
AIC	3508.595

5.0 CONCLUSION

The Poisson Regression Model, GPR, and NBR were conducted to determine the better model to use in modeling data on number of road traffic crashes within the Anambra state Command of Federal Road Safety Commission, Nigeria. The criterion for selection of the best model used is the AIC. The best model is that with the smallest AIC value. This happened to be the NBR model. The deviance for the Poisson regression model **(Table C)** is larger than the deviance of the Negative Binomial Regression and the Generalized Poisson Regression models **(Tables E and G)** thus indicating the existence of significant over-dispersion if the Poisson Regression Model were adopted. To test for over-dispersion, AIC of Poisson against Negative Binomial Regression model and Generalized Poisson Regression are obtained and compared. On the basis of the AIC values in tables **C, E** and **G** the estimated AIC for GPR model **(Table G)** is 3508.595 whereas it is 6266 for the PR **(Table C)** and 2742 for the NBR model **(Table E)**. The smallest AIC value is that of the negative binomial regression model. Therefore, the best model for the number of road traffic crashes in Anambra state road safety command is best modeled and described using the negative binomial regression model.

REFERENCES

[1] Bozdogan, H. (2000). “Akaike's Information Criterion and Recent Developments in Information Complexity”. Mathematical Psychology, 44 , 62-91.

[2] Cameron, A.C and Trivedi, P. K (1998), “Regression analysis of count data.” Cambridge University press Cambridge, UK.

[3] Consul P. C. and Famoye F. (1992), “Generalized Poisson regression model”, communications in statistics (theory and methodology) vol. 2, no.1, 89-109.

[4] Famoye F, John T. W. and Karan P. S. (2004), “On the generalized Poisson regression model with an application to accident data”. Journal of data science 2 (2004), 287-295

[5] Greene, W (2008) “Functional Forms For The Negative Binomial Model For Count Data”. Foundations and Trends in Econometrics. Working Paper, Department of Economics, Stern School of Business, New York University. 585-590.

[6] Moshe B. H. and Haim L. (1981): “Statistics: Decisions and Applications in Business and Economics”. Random House Inc. New York.

[7] Nantulya, V.M and Reich M.R (2002), “The neglected epidemic: Road traffic injuries in developing countries”. Br. Med. Journal, 324:1139-1141

[8] O’Neill B. and Mohan D. (2002).”Reducing motor vehicle crash deaths and injuries in newly motorizing countries”. BMJ., 324:1142-1145

[9] Oskam J., Kingma J and Klasen H.J (1994), “The Groningen trauma study. Injury patterns in a Dutch trauma centre”. Eur. J. Emergency Med., 1:167-172.