ISO 9001 - 2015

ISSN 2349 - 4891

Monthly

IJRRAS

IF 4.665

Volume 4,Issue 3, March 2017

International Journal of Recent Research and Applied Studies

SURRAGH PUBLICATIONS

Original Article

ISSN: 2349 - 4891

International
Journal of Recent Research and Applied Studies
(Multidisciplinary Open Access Refereed e-Journal)

The Application of Generalized Linear Regression Models Cope with the Types of Vehicles in Road Accidents

Philip Arokiadoss. A[1], Muthu. C[2] & Arulanandu. U[3]

[1]*Doctoral Research Scholar, Research Department of Statistics, St. Joseph's College, Tiruchirappalli, Tamil Nadu, India. & Assistant Professor, Department of Mathematics, Thanthai Hans Roever College, Perambalur, Tamil Nadu, India.*
[2]*Associate Professor, Research Department of Statistics, St. Joseph's College, Tiruchirappalli, Tamil Nadu, India.*
[3]*Professor of Statistics, Anbil Dharmalingam Agricultural College &Research Institute, Tiruchirappalli, Tamil Nadu, India.*

Received 5th January 2017, Accepted 1st March 2017

Abstract

In the modern era, Road accident prediction models are priceless tools that have valuable applications in road accidents protective analysis. This paper focuses on the Generalized Linear Regression (GLR) modeling on the number of people died, injured with primarily involved vehicles by road accidents for the years 2001-2015 in Tiruchirappalli District and the exhaustive analysis of the data using two statistical techniques such as Poisson regression and Negative Binomial regression to fit a model to the data. To propose improvement measures to prevent road accidents and to derive a model for accident parameters. This paper suggests procedures for developing prudent models, i.e. models that are not overfitted, and best-fit models. The respective models were used to identify the vehicles that caused more people died and injured.

Keywords: GLR, Negative Binomial Regression Models, Number of people died, Number of People injured, Poisson Regression Model and Vehicles.

Introduction

Transportation is a non-separable part of any society. More than six thousand years, the constructions of roadways have been doing by human beings. Early roadways carried pedestrians, hoofed animals and simple wheeled vehicles such as wagons. As traffic volumes grew, it became apparent that traffic regulation would be necessary for safety and efficiency. In the first century BC, wheeled traffic was banned by Julius Caesar within the ancient Rome during the certain hours of the day (Swaha Bhattacharya, 2008). After that, Ms. Mary Ward was the first documented victim of vehicular accident that took place on August 31, 1869, (Akarro RJ, 2009) the worldwide road accidents fatalities have been raised to nearby 1.2 million/year. Almost three quarter of deaths resulting from motor vehicle crashes occur in developing country (Odero W et., 1997). In India, the motor vehicle population has increased significantly over the last four decades in a faster rate than the economic and population growth. The surge in motorization coupled with expansion of the road network has brought with it the challenge of addressing adverse factors such as the increase in road accidents. It is consequently the need of the hour to take suitable preventive measures, so that loss of precious lives can be brought down to a minimum.
There are plentiful techniques available for estimating the number of accidents. The Multiple regression technique was helped to find out the prior relationship between the number of accidents, human population and vehicle population (Philip Arokiadoss et al, 2016). Hadi et al, (1995) and Anis (1996) have been made to describe the discrete count road accident data and to produce more accurate and reliable models through the use of Generalized Linear Models (GLM) with Poisson and negative binomial distributions.

The Poisson regression model is also a technique of Generalized Liner model which is used to describe count data as a function of a set of predictor variables. To investigate the incidence and mortality of chronic diseases, it has been extensively used both in human and in veterinary epidemiological studies in the last two decades. Also, Poisson regression has been applied in the analysis of accident data for modelling Road accidents in different parts of the world. Among its numerous applications, Poisson regression has been mainly applied to compare exposed and unexposed cohorts and to evaluate the causes of road traffic accidents. In recent scenario, Poisson type regression models have been used to model count response variable affected by one or more covariates. The generalized event count models based on the Poisson, negative binomial distributions developed by King (1989). Winkelmann and Zimmermann (1994) 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. In order to address the issue of over-dispersion Abdel-Aty and Radwan (2000)

Correspondence
Philip Arokiadoss. A
E-mail: mail2philip.a@gmail.com, Ph. +9194879 09159

and Guevara et al., (2004) used negative binomial distribution which allows variance to exceed the mean. The objective of the paper is to evaluate primarily involved vehicles and its relation between accident per year and inclusive Parameters such as number of person died and injured and to develop Generalized Linear Regression models with test their validity.

Methodology

The corresponding Road accident data is to be collected from District Crime Records Bureau (DCRB), District Police Office, Tiruchirappalli district. It includes no. of vehicles involved, number of person died, injured, and etc., the data was taken from 2001 to 2015. Also, the accidents data are represented as figures and characterized the road accidents by types of vehicles are given in years ancient. The analysis of the data would be done by using MS-Excel and RStudio software. The proposed methodology has shown in the Figure (1),

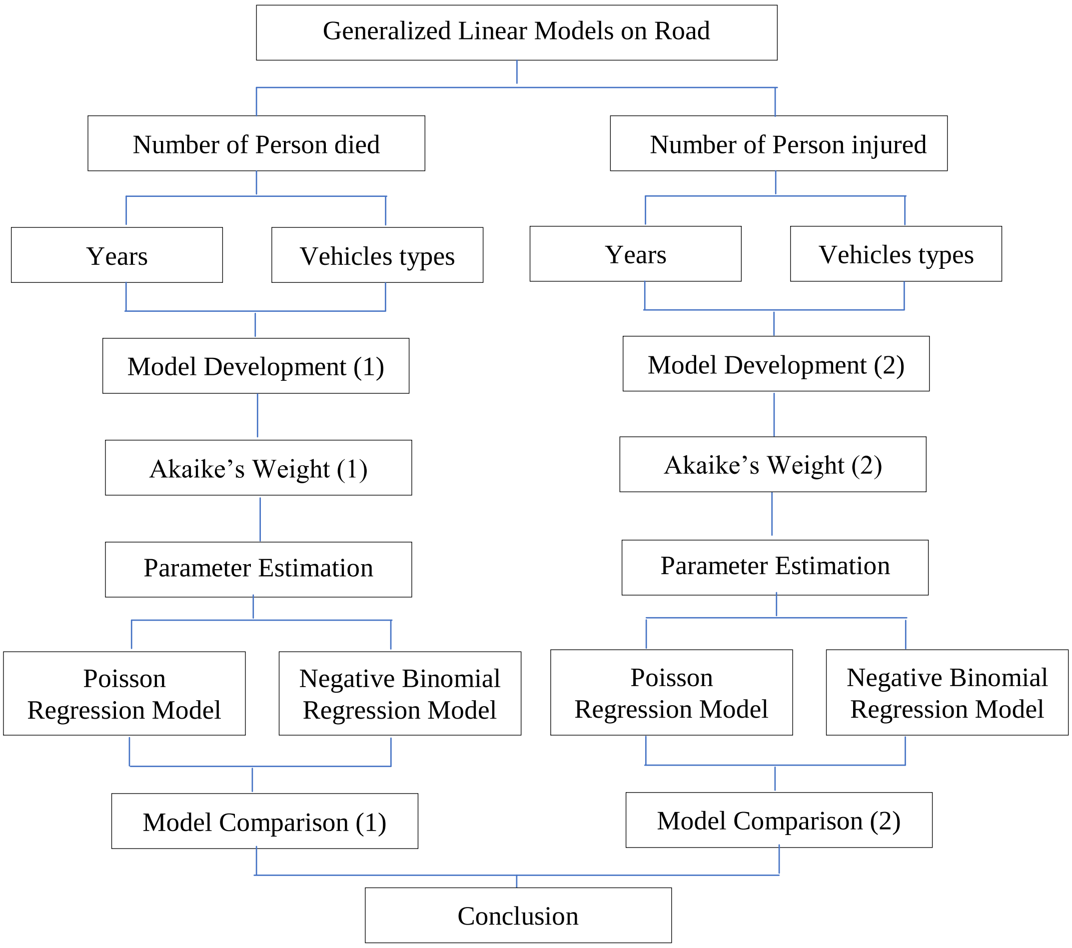

Figure I
Organization chart

The Generalized Linear Model (GLM) is an extension of the Linear Model to include response variables that follow any probability distribution in the exponential family of distributions. Many commonly used distributions in the exponential family are the normal, binomial, Poisson, exponential, gamma and inverse Gaussian distributions. GLM was first introduced by Nelder and Wedderburn (1972). It provided a unified framework to study various regression models, rather than a separate study for each individual regression. GLM consists of three components:

a. *Random Component* – refers to the probability distribution of the response variable (Y) given the explanatory variable X_{ij}.
b. *Systematic Component* - specifies the explanatory variables ($X_1, X_2, \ldots X_m$) in the model, more specifically their linear combination in creating the so called *linear predictor*.
   $$i.e.,\ Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_m X_{mi} \quad \text{--- (2.1)}$$
c. *Link Function*, η_i *or* $g(\mu_i)$ - specifies the link between random and systematic components. It says how the expected value of the response relates to the linear predictor of explanatory variables;
   $$\eta_i = g(E(Y_i)) = E(Y_i) \quad \text{--- (2.2)}$$

GLM is an extension of the classical linear models. It includes linear regression models, analysis of variance models, Logistic regression models, Poisson regression models, Zero-inflated Poisson regression models, Negative Binomial regression models, log-linear models, as well as many other models. The above models share a number of unique properties, such as linearity and a common method for parameter estimation. In spite of its recent wide application, Poisson regression model remains partly poorly known, especially if compared with other regression techniques, like linear, logistic and Cox regression models. The Poisson regression model assumes that the sample of n observations, are observations on independent Poisson variables Y_i with mean $E(Y_i)=\lambda_i$. If this model is correct, the equal variance assumption of classic linear regression is violated, since the Yi have means equal to their variances. So, we fit the generalized linear model,

$$\text{Log}(\lambda i) = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_m X_{mi} + \varepsilon_{ij} \qquad \text{-- (2.3)}$$

The major assumption of Poisson model is,

$$E\,(y_i\; x_i) = Var\,(y_i\; x_i) \qquad \text{--- (2.4)}$$

If $Var\ (y_i\ x_i) > E\ (y_i\ x_i)$ then there is over-dispersion. If, $Var\ (y_i\ x_i) < E\ (y_i\ x_i)$ then under-dispersion has occurred. The Poisson model does not allow for over or under dispersion. A comfortable model is obtained by using the Negative binomial distribution instead of the Poisson distribution. To estimate the parameters using the Maximum Likelihood Estimation (MLE). Negative binomial regression is a prevalent generalization of Poisson regression because it slackens the highly restrictive assumption that the variance is equal to the mean made by the Poisson model. A measure of discrepancy between observed and fitted values is the deviance. Deviance is defined as the log likelihood of the final model, multiplied by (-2). The formula to estimate the deviance has given below.

$$D(\hat{y}_i) = \sum_{i=1}^{k} 2\left[y_i \log \tfrac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i) \right] \qquad \text{--- (2.5)}$$

The Pearson Chi-square is used to measure the log-likelihood value to measure the goodness of fit that compares the predicted values of the dependent variable with all other observed values.

$$\chi^2 = \sum_{i=1}^{k} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i} \;;$$

where $\hat{y}_i$ be the predicted value of y_i --- (2.6)

The consequence of this test for comparing models mentioned is usually equivalent to evaluating the dispersion parameter. Akaike Information Criteria (AIC) or Bayesian Information Criteria (BIC) is applied. AIC and BIC are calculated as:

AIC = -2LL+2k; BIC =-2LL+k log n --- (2.7)

Where LL be a log-likelihood

k – number of parameter

n – number of observations

According to Liu, W. and Cela, J. (2008) renowned that the least AIC is the perfect model fit. Consequence of model evaluation AIC and BIC are similar and their values are close together. Meanwhile, the Pearson Chi-square test, deviance, likelihood ratio test, AIC and BIC are very acquainted to those who used the General Linear Models (Noriszura Ismail et al, 2007) with Poisson error structure for claim frequency analysis. Since these measures might be implemented to the Poisson and Negative Binomial Regression Models as well.

Analysis and Interpretation

Table 1

Types of vehicles primarily responsible of Road accidents (2001-2015)

Parameter		Type of Vehicle Primarily Responsible										
		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
FA	Nos.	1444	189	727	6	270	760	955	424	0	94	458
	Mean	96.3	12.6	48.5	0.4	18.0	50.7	63.7	28.3	0.0	6.3	30.5
	%	27.1	3.5	13.6	0.1	5.1	14.3	17.9	8.0	0.0	1.8	8.6
NFA	Nos.	4211	257	330	38	669	3161	3039	228	0	383	1534
	Mean	280.7	17.1	22.0	2.5	44.6	210.7	202.6	15.2	0.0	25.5	102.3
	%	30.4	1.9	2.4	0.3	4.8	22.8	21.9	1.6	0.0	2.8	11.1
PD	Nos.	1738	194	845	14	322	916	1134	361	0	100	522
	Mean	115.9	12.9	56.3	0.9	21.5	61.1	75.6	24.1	0.0	6.7	34.8
	%	28.3	3.2	13.7	0.2	5.2	14.9	18.5	5.9	0.0	1.6	8.5
PI	Nos.	7125	750	3818	34	1178	6370	4389	1988	0	492	3371
	Mean	475.0	50.0	254.5	2.3	78.5	424.7	292.6	132.5	0.0	32.8	224.7
	%	24.1	2.5	12.9	0.1	4.0	21.6	14.9	6.7	0.0	1.7	11.4

*X_1 = Two-wheelers; X_2 = Three-wheelers; X_3 = Car; X_4 = Jeep; X_5 = Taxi; X_6 = Bus; X_7 = Truck; X_8 = Tempo; X_9 = Articulated Vehicles; X_{10} = Tractor; X_{11} = Others / Unknown Vehicles

People died by various vehicles

The experimental results reveal that Road accident data from 2001 to 2015 which involved two-wheelers constitute 28.3% of people died and it was tabulated in the Table 1. This was followed closely by Truck which died 1,134 representing 18.5% of those who were died by road accidents.

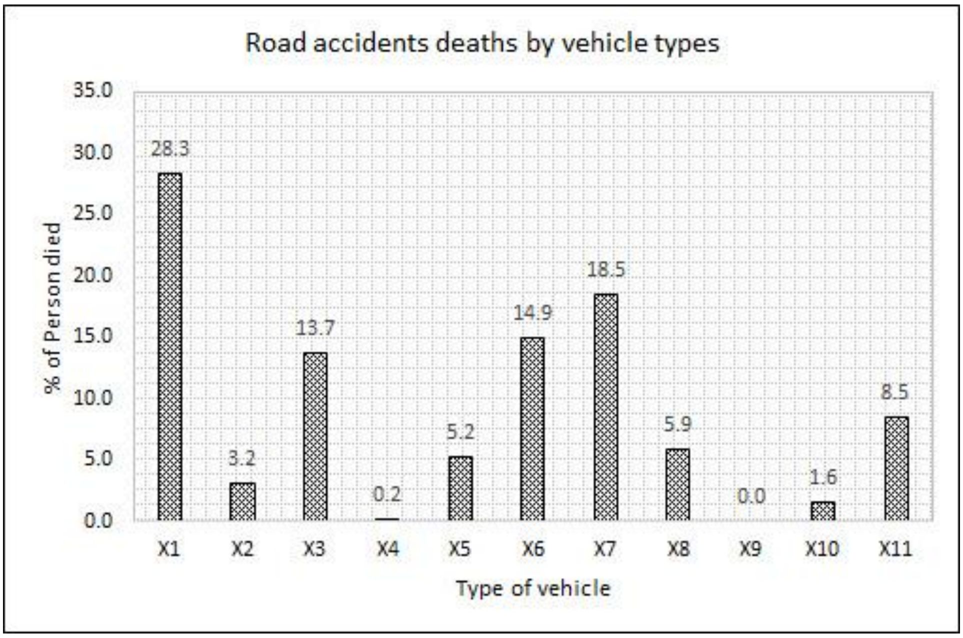

Figure II
Percentage of people died by various types of vehicles

Buses (14.9%) were the third on the list of type of vehicles which die most people in accidents with 916 people who were died which graphically illustrated in the Figure (2). Hostilely, the type of vehicle which no one died nor injured by Articulated Vehicles such as bulldozer, tipper, mixer and loading box. In order to introduce the Poisson distribution and consider the person died by the road accident data for modeling the linear function over the same period. The result is a Generalized Linear Regression Model (GLRM) with Poisson response and log link function.

Akaike's Weight

Model	Specification	AIC	Δ_i AIC	RL	W_i (AIC)
(a)	$log(\lambda_a)_{pd} = \alpha_a + \beta_i\ v\_type;\ i=1, 2, 3, ..., 11.$	1182.74	207.90	0.00	0.00
(b)	$log(\lambda_b)_{pd} = \alpha_b + \beta_i\ year;\ i=1, 2, 3, ..., 15.$	6330.35	5355.51	0.00	0.00
(c)	$log(\lambda_c)_{pd} = \alpha_c + \beta_i\ v\_type + \beta_j\ year;\ i=1, 2, 3, ..., 11; j=1, 2, 3, ..., 15.$	974.84	0.00	1.00	1.00

** pd=person died; v_type=vehicle types; RL=Relative Likelihood;*

The above Akaike's Weight specifies that the perfect model (approx. 99%) which fit the types of vehicles that people died in Tiruchirappalli district from 2001-2015, is Model (c). Seemingly, it doesn't any entrant, owing to Model (c) has the smallest AIC, correspondingly it has the greatest cost of Akaike weight (Wi).

Goodness of fit

Data Criteria	Value	df	Value/df
Deviance	200.155	140	1.430
Pearson Chi-Square	221.404	140	1.581
AIC	979.842		
BIC	1057.491		

The goodness-of-fit can be based on the deviance statistic that is defined by Famoye (1993). The deviance statistic can be approximated by a chi-square distribution when μ_i's are large. We use the log-likelihood value to measure the goodness-of-fit of the regression models. A measure of discrepancy between observed and fitted values is the deviance. We evaluate the deviance (200.155) as Chi-square distributed with the model degrees of freedom (140). Hence, we conclude that the model fit reasonably well because the compilation of likelihood ratio chi-squared test is no statistically significant with degrees of freedom (140) against the intercept.

Table 2

Estimation results for Poisson Regression and Negative Binomial Regression Models- People died by Road accidents

Parameter	Model 1: Poisson Regression			Model 2: Negative Binomial Regression		
	Estimate	t-statistic	Pr(>|t|)	Estimate	t-statistic	Pr(>|t|)
(Intercept)	4.23735	50.963	< 2e-16 ***	4.23700	64.058	< 2e-16 ***
V_typeX2	-2.15571	-22.603	< 2e-16 ***	-2.15600	-28.418	< 2e-16 ***
V_typeX3	-0.68423	-12.895	< 2e-16 ***	-0.68420	-16.203	< 2e-16 ***
V_typeX4	-4.78451	-14.176	< 2e-16 ***	-4.78400	-17.827	< 2e-16 ***
V_typeX5	-1.64902	-21.550	< 2e-16 ***	-1.64900	-27.090	< 2e-16 ***
V_typeX6	-0.60355	-11.679	< 2e-16 ***	-0.60350	-14.675	< 2e-16 ***
V_typeX7	-0.39006	-8.066	2.94e-13 ***	-0.39010	-10.134	< 2e-16 ***
V_typeX8	1.37383	-20.095	< 2e-16 ***	-1.37400	-25.259	< 2e-16 ***
V_typeX9	-21.99906	-0.020	0.984296	-41.74000	0.000	1.00000
V_typeX10	-2.81840	-21.771	< 2e-16 ***	-2.81800	-27.375	< 2e-16 ***
V_typeX11	-1.16590	-18.495	< 2e-16 ***	-1.16600	-23.246	< 2e-16 ***
YearY2	0.06847	0.624	0.533665	0.06854	0.785	0.43233
YearY3	0.24032	2.279	0.024198 *	0.24040	2.865	0.00417 **
YearY4	0.32913	3.181	0.001808 **	0.32920	3.999	6.37e-05 ***
YearY5	0.42109	4.147	5.82e-05 ***	0.42110	5.212	1.87e-07 ***
YearY6	0.40284	3.952	0.000122 ***	0.40280	4.968	6.76e-07 ***
YearY7	0.48853	4.874	2.92e-06 ***	0.48850	6.126	9.01e-10 ***
YearY8	0.49815	4.979	1.85e-06 ***	0.49810	6.258	3.89e-10 ***
YearY9	0.62176	6.356	2.73e-09 ***	0.62180	7.989	1.36e-15 ***
YearY10	0.59172	6.017	1.48e-08 ***	0.59180	7.563	3.94e-14 ***
YearY11	0.61327	6.260	4.42e-09 ***	0.61330	7.868	3.61e-15 ***
YearY12	0.61327	6.260	4.42e-09 ***	0.61330	7.868	3.61e-15 ***
YearY13	0.64887	6.664	5.65e-10 ***	0.64890	8.376	< 2e-16 ***
YearY14	0.67527	6.967	1.16e-10 ***	0.67520	8.755	< 2e-16 ***
YearY15	0.65707	6.758	3.47e-10 ***	0.65700	8.493	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The above Table 2 reveals that the model-1 estimated by the Poisson regression. The AIC of this model was 979.84; the null deviance was 5801.30 on 164 degrees of freedom and residual deviance of 200.15 on 140 degrees of freedom following the chi-square distribution (χ^2) with one degree of freedom. However, an assumption of Poisson distribution which is the equality of the mean and variance which means that the dispersion parameter should always be closer to 1 has been violated. The dispersion parameter (1.58146) of the above model is far greater than 1, an indication of over dispersion in the data. The parameters of the model have been over estimated and will not give a true reflection of number of people likely to be died through road accidents in a given types of vehicles for a particular year. To eliminate this error by model-2, Negative Binomial regression which was used to validate the model that the parameter estimates reduced and the standard errors also decreased. The parametric analysis for the comparison between the Poisson and Negative Binomial regressions for evaluating the best fitted model using Table (3) as given below.

Table 3
Results of Model Evaluations and Comparisons

Measures	Poisson Regression Model	Negative Binomial Regression Model
Null deviance	5801.30	5797.24
Degrees of freedom	164.00	164.00
Residual deviance	200.15	200.05
Degrees of freedom	140.00	140.00
AIC	979.84	981.86
Dispersion parameter	1.58	1.00
No. of Fisher Scoring	15.00	1.00

We observed from the above Table 3, the Poisson Regression model is actually the best model which fit the model for Number of People who were died by Different Types of Vehicles data from 2001-2015. Because AIC of the Poisson model (979.84) is less than the Negative Binomial model (981.86). As the dispersion parameter, has reduced from 1.58 which was giving by the Poisson regression model to 1.00. Table 2 reveals that the first group (β_i) in a data as the base level by default and such as V_typeX1 (Two wheelers) and the year 2001 were selected as the base levels for comparison in the analysis of the parameter estimates in the Poisson regression model. The intercept was found to be 4.2374 which was very significant at 95% significant level with p-value of < 2e-16. Excluding Articulated Vehicles (V-typeX9) which was not significantly different from the two-wheelers in the model, the rest of the vehicles were all significantly smaller than the base level in the model at 5% α-level for every year. For instance, V-typeX2 (three wheelers) was found to have parameter estimate of -2.156 less than the logarithm of the expected number of people who were died by two-wheelers for every year. Also, the expected number of people who were died by others/unknown vehicles (V_typeX11) was $e^{-1.166}$ = 0.311642 times less than that of two-wheelers for every year. The Table 2, further exposes that the second group (β_j) such that the expected number of people who were died by different types of vehicles for the years 2002, 2003 and 2004 were not significantly different from 2001 for all types of vehicles in the model giving that the year 2001 is the base level. 2010 was found to be the year which had most people died by road accident for all types of vehicles in Tiruchirappalli. It was found that 2015 had $e^{0.6571}$=1.929132 times more than the expected number of people died in 2001 for all types of vehicles in Tiruchirappalli District. The Fitted model is existing in equation 3(a) below.

$$log(\lambda_c)_{pd} = 4.2374 - 2.156(X_2) -0.684(X_3) -4.785(X_4) -1.649(X_5) -0.604(X_6) -0.390(X_7) +1.3738 (X_8) -22.9991(X_9) -2.818(X_{10}) -1.166(X_{11}) +0.0685(Y_2) +0.2403(Y_3) +0.3291(Y_4) +0.4211(Y_5) +0.4028(Y_6) +0.4885(Y_7) +0.4982(Y_8) +0.6218(Y_9) +0.5917(Y_{10}) +0.6133(Y_{11}) +0.6133(Y_{12}) +0.6489(Y_{13}) +0.6753(Y_{14}) +0.6571(Y_{15}) \quad \text{---------} \ 3(a)$$

where X_1, X_2…, X_{11} represent the primarily involved vehicle types who there is the number of people died by road accidents and Y_1, Y_2…, Y_{15} denotes the year 2001, 2003, …, 2015 respectively.

People injured by various vehicles

The Table 1 depicts that two-wheelers were responsed maximum (24.1%) number of people injured. This was followed nearly by Buses which injured 6,370 representing 21.6% of those who were injured by road accidents. Trucks (14.9%) were the third on the list of type of vehicles which kill most people in accidents with 4,389 people who were injured for the period 2001 to 2015. The type of vehicle which died and injured the least number of people for the years under consideration is Jeep. The percentage of people injured by various types of vehicles in Tiruchirappalli district is as shown in the Figure III below.

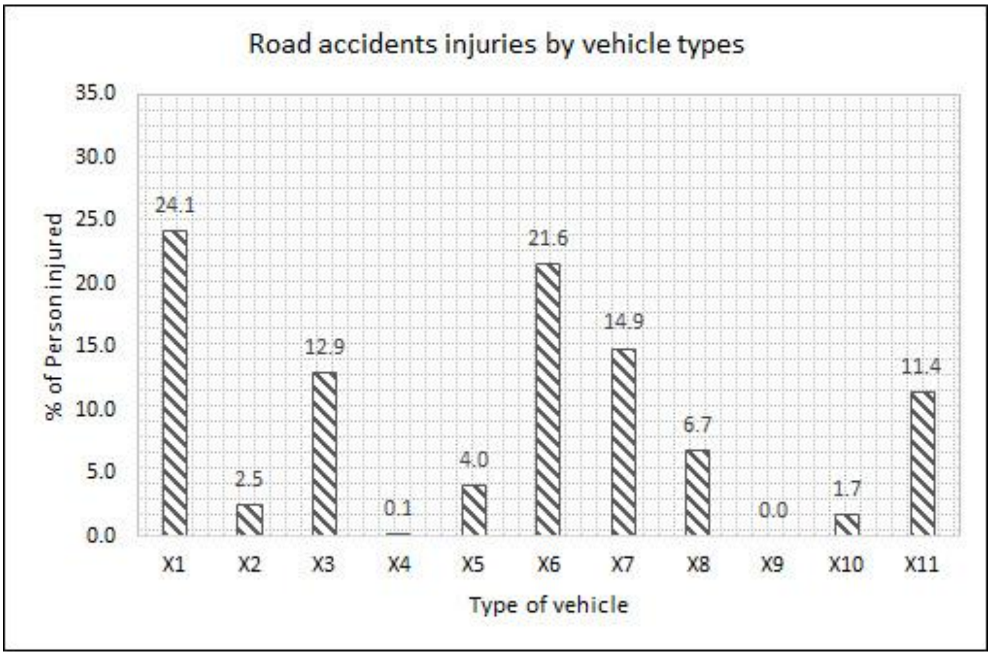

Figure III
Percentage of people injured by various types of vehicles

The Figure III represents the percentage of people injured through road accidents in Tiruchirappalli district from 2001 to 2015. In order to the Poisson distribution and consider the number of person injured by the road accident data for modeling the linear function over the period 2001-2015. The following result is a GLR model with Poisson response and log link function.

Akaike's Weight

Model	Specification	AIC	Δ_i AIC	RL	W_i (AIC)
(d)	$log(\lambda_d)_{pi} = \alpha_d + \beta_i\ v\_type;\ i=1, 2, 3, \ldots\ldots, 11.$	3178.10	2506.06	0.00	0.00
(e)	$log(\lambda_e)_{pi} = \alpha_e + \beta_i\ year;\ i=1, 2, 3, \ldots\ldots, 15.$	28815.00	28142.96	0.00	0.00
(f)	$log(\lambda_f)_{pi} = \alpha_f + \beta_i\ v\_type + \beta_j\ year;\ i=1, 2, 3, \ldots., 11;\ j=1, 2, 3, \ldots., 15.$	2389.50	1717.46	1.00	1.00

* *pi=person injured; v_type=vehicle types; RL=Relative Likelihood;*

The Akaike's Weight prefers that the Model (f) has perfectly fit (approx. 99.9%) for the types of vehicles that people injured in Tiruchirappalli district from 2001-2015. Apparently, it doesn't any competitor value, since it has the smallest AIC, likewise it has the extreme cost of Akaike weight (W*i*).

Goodness of fit

Data Criteria	Value	df	Value/df
Deviance	1381.814	140	9.870
Pearson Chi-Square	1341.345	140	9.581
AIC	2389.503		
BIC	2467.182		

The goodness-of-fit can be based on the deviance statistic. It can be move toward by a chi-square distribution when μ_i's are large. We use the log-likelihood value to measure the goodness-of-fit of the regression models. A measure of discrepancy between observed and fitted values is the deviance. We evaluate the deviance (1381.814) as Chi-square distributed with the model degrees of freedom (140). Hence, we conclude that the model fit reasonably well because the compilation of likelihood ratio chi-squared test is no statistically significant with degrees of freedom (140) against the intercept.

Table 4

Estimation results for Poisson Regression and Negative Binomial Regression Models- People died by Road accidents

Parameter	Model 1: Poisson Regression			Model 2: Negative Binomial Regression		
	Estimate	t-statistic	p-value	Estimate	t-statistic	p-value
(Intercept)	5.809	65.093	< 2e-16 ***	5.831	52.372	< 2e-16 ***
V_typeX2	-2.673	-18.525	< 2e-16 ***	-2.618	-25.116	< 2e-16 ***
V_typeX3	-0.748	-11.566	< 2e-16 ***	-0.702	-7.328	2.34e-13 ***
V_typeX4	-2.251	-18.946	< 2e-16 ***	-2.234	-22.076	< 2e-16 ***
V_typeX5	-0.624	-10.049	< 2e-16 ***	-0.641	-6.692	2.21e-11 ***
V_typeX6	-5.234	-10.396	< 2e-16 ***	-5.208	-27.851	< 2e-16 ***
V_typeX7	-1.803	-18.495	< 2e-16 ***	-1.805	-18.257	< 2e-16 ***
V_typeX8	-0.112	-2.099	0.03765 *	-0.085	-0.890	0.37355
V_typeX9	-0.485	-8.157	1.76e-13 ***	-0.487	-5.097	3.46e-07 ***
V_typeX10	-1.276	-16.258	< 2e-16 ***	-1.288	-13.276	< 2e-16 ***
V_typeX11	-21.454	-0.021	0.98314	-38.440	0.000	0.99998
YearY2	0.010	0.086	0.93169	0.048	0.371	0.71097
YearY3	0.168	1.487	0.13918	0.223	1.731	0.08341
YearY4	0.310	2.828	0.00537 **	0.314	2.437	0.01480 *
YearY5	0.228	2.045	0.04273 *	0.226	1.750	0.08015
YearY6	0.224	2.002	0.04727 *	0.193	1.494	0.13527
YearY7	0.354	3.256	0.00142 **	0.310	2.408	0.01606 *
YearY8	0.352	3.236	0.00151 **	0.320	2.485	0.01294 *
YearY9	0.442	4.134	6.11e-05 ***	0.419	3.265	0.00109 **
YearY10	0.489	4.624	8.49e-06 ***	0.435	3.389	0.000703 ***
YearY11	0.484	4.564	1.09e-05 ***	0.429	3.341	0.000834 ***
YearY12	0.470	4.425	1.93e-05 ***	0.393	3.062	0.002196 **
YearY13	0.509	4.829	3.56e-06 ***	0.416	3.243	0.001184 **
YearY14	0.534	5.084	1.16e-06 ***	0.462	3.605	0.000312 ***
YearY15	0.531	5.053	1.33e-06 ***	0.452	3.522	0.000429 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The Table 4 shows that the model-1 valued by the Poisson Regression model. Still, one assumption of Poisson distribution which is the equality of the mean and variance which means that the dispersion parameter should always be closer to 1 has been violated. The dispersion parameter (9.581038) of the above model is far greater than 1, an indication of over dispersion in the data. To eliminate such an error by model-2, Negative Binomial regression which was used to validate the model that the parameter estimates reduced and the standard errors also decreased. The parametric analysis for the comparison between the Poisson and Negative Binomial regression for goodness of fit test of the model is shown in Table below.

Table 5

Results of Model Evaluations and Comparisons

Measures	Poisson Regression Model	Negative Binomial Regression Model
Null deviance	28643.80	3629.00
Degrees of freedom	164.00	164.00
Residual deviance	1381.80	176.84
Degrees of freedom	140.00	140.00
AIC	2389.50	1517.30
Dispersion parameter	9.58	1.00
No. of Fisher Scoring	13.00	1.00

Table 5 found that, the Negative Binomial Regression model is actually the best model which fit the model for Number of People who were injured by various types of vehicles for Road accident data from 2001-2015. Because AIC of the Negative Binomial model (1517.30) is less than the Poisson model (2389.50). As the dispersion parameter, has reduced from 9.58 which was giving by the Poisson regression model to 1.00. From the Table 4, R Studio takes the model-2 in a data as the base level by default as two-wheelers (V_typeX1) and the year 2001 were selected as the base levels for comparison in the analysis of the parameter estimates in the negative binomial regression model. The intercept was found to be 5.809 which was very significant at 95% significant level with p-value of < 2e-16 following a Pearson Chi-square distribution 1341.345 with 140 degrees of freedom others/unknown vehicles (V-typeX11) which was not significantly different from the two-wheelers in the model, the rest of the vehicles were all significantly smaller than the base level in the model at 5% α-level for every year. For instance, Tempo (V-typeX8) was found to have parameter estimate of -0.085 less than the logarithm of the expected number of people who were injured by two-wheelers for every year. The evident also point to Table 4 that the expected number of people who were injured by Buses (V_typeX6) was $e^{-5.208}$ =0.0055 times less than that of two wheelers for every year. Further the Table (4) reveals that the expected number of people who were injured by different types of vehicles for the years 2002, 2003, 2005 and 2006 were not significantly different from 2001 for all types of vehicles in the Negative Binomial Regression model. In the year of 2014 was found to be the maximum number of people injured by road accident for all types of vehicles in Tiruchirappalli District. It was found that 2010 had $e^{-0.462}$ = 0.6300 times more than the expected number of people injured in 2001 for all types of vehicles in the district. The Fitted model is existing in equation as written below.

$$log(\lambda_f)_{pi} = 5.831 - 2.618(X_2) -0.702(X_3) -2.234(X_4) -0.641(X_5) -5.208(X_6) -1.805(X_7) +0.085 (X_8) -0.487(X_9) -1.288(X_{10}) -38.440(X_{11}) +0.048(Y_2) +0.223(Y_3) +0.314(Y_4) +0.226(Y_5) +0.193(Y_6) +0.310(Y_7) +0.320(Y_8) +0.419(Y_9) +0.435(Y_{10}) +0.429(Y_{11}) +0.393(Y_{12}) +0.416(Y_{13}) +0.462(Y_{14}) +0.452(Y_{15}) \quad \text{------- 3(b)}$$

where X_1, X_2…, X_{11} represent the primary involved vehicle types who there is the number of people injured by road accidents and Y_1, Y_2…, Y_{15} denotes the year 2001, 2002, …, 2015 in that order.

Conclusion

The outcome of the paper accomplish that the Poisson regression model is fitted perfectly for the number of people who were died by various types of vehicles. It was identified that two-wheelers and Trucks were primarily involved and triggered more people to die in the road accidents during 2001-2015. Also, the result proposed of 2014 was found to be the year which had most people died by road accident. Furthermore, negative binomial regression model quite better to fit the number of people who were injured by various types of vehicles. It was observed from the model two-wheelers and Buses were chiefly involved and triggered more people to injure by the road accidents. It was also found that 2014 to be the year which had utmost people injured by the road accidents. In spite of the ever-increasing number of people using our road and the increasing occurrence of road accidents in Tiruchirappalli district, the problem can be reduced if every citizen of the district can strictly keep to all the preventive measures. In the meantime, the type of vehicle involved in the accident affects the number of people expected to be died / injured; drivers of vehicles such as Two-wheelers, Trucks and Buses should be given special training to be able to avoid preventable accidents. Road accidents injuries are predictable and preventable, but good data are important to understand the ways in which road safety interventions and technology can be successfully transferred from developed countries where they have proven effective. Awareness of the consequences of road accident injuries are lagging among riders, drivers and the public.

References

1. Abdel-Aty, M and Radwan, AE, (2000). 'Modelling Traffic Accident Occurrence and Involvement', Accident Analysis and Prevention; 32: issue (5): pp 633-642.
2. Akarro RJ (2009). People's Opinions on the Causes of Road Accidents in Selected Places of Tanzania. European Journal of Social Sciences; 9: 615-626.
3. Anis, G. (1996). An Application of Generalized Linear Modelling to the Analysis of Traffic Accidents, Traffic Engineering Control; 37: issue (12): pp 691-696.
4. Famoye, F. (1993). Restricted generalized Poisson regression model. Communications in Statistics – Theory and Methods; 22: 1335-1354.
5. Guevara, FL., Washington SP., (2004). Forecasting Crashes at the Planning Level Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona, Transportation Research Record: Journal of the Transportation Research Board, No.1897, TRB, National Research Council, Washington, D.C., pp 491-499.
6. Hadi, MA., Aruldhas, J., Chow, LF., Wattleworth, JA., (1995). Estimating Safety Effects of Cross-section Design for Various Highway Types using Negative Binomial Regression Transportation Research Record: Journal of the Transportation Research Board, No. 1500, TRB National Research Council, Washington, DC; pp.169-177.
7. King, G. (1989). Variance specification in event count models: From restrictive assumptions to a

generalized estimator. American Journal of Political Science; 33: 762-784.
8. Liu, W. and Cela, J. (2008). Count Data Models in SAS. SAS Global Forum 2008: Statistics and Data Analysis. Paper 371-2008.
9. Noriszura Ismail, Abdul Aziz Jemain (2007). Handling Over dispersion with Negative Binomial and Generalized Poisson Regression Models; Casualty Actuarial Society Forum; 103-158.
10. Odero W, Garner P, Zwi A. (1997). Road traffic Injuries in developing countries: A comparative review of epidemiological studies: Trop. Med Int Health; 2: 445-60.
11. Philip Arokiadoss A, Muthu C, Arulanandu U and Joseph M. (2016). Exploring the relationship between human and vehicle population – It's perspective study in Road accidents, International Journal of Mathematics and Computer Applications Research, 6: 23-34.
12. Swaha Bhattacharya (2008). Environmental Psychology, Dr. N.K. Singh for Global Vision Publishing House, New Delhi, India.
13. Winkelmann, R. and Zimmermann, K. F. (1994). Count data models for demographic data. Mathematical Population Studies; 4: 205-221.