

RESEARCH

Open Access



Analysis of injuries and deaths from road traffic accidents in Iran: bivariate regression approach

Soodeh Shahsavari¹, Ali Mohammadi¹, Shayan Mostafaei^{2,3}, Ehsan Zeresghi², Seyyed Mohammad Tabatabaei⁴, Mohsen Zhaleh⁵, Meisam Shahsavari⁶ and Frouzan Zeini^{2*}

Abstract

Backgrounds: This study aims to estimate and compare the parameters of some univariate and bivariate count models to identify the factors affecting the number of mortality and the number of injured in road accidents.

Methods: The accident data used in this study are related to Kermanshah province in march2020 to march2021. Accidents areas were divided into 125 areas based on density characteristics. In a one-year period, 3090 accidents happened on the suburban roads of Kermanshah province, which resulted in 398 deaths and 4805 injuries. Accident information, including longitude and latitude of accident location, type of accident (fatal and injury), number of deaths, number of injuries, accident type, the reason of the accident, and the kind of accident were all included as population-level variables in the regression models. We investigated four frequently used bivariate count regression models for accident data in the literature.

Results: In bivariate analysis, except for the DNM model, there is a reasonable decrease in the AIC measures of the saturated model compared to the reduced model for the other three models. For the injury models, MSE is lowest, respectively for DIBP (137.87), BNB (289.46), BP (412.36) and DNM (3640.89) models. These results are also established for death models. But, in univariate analysis, only injury models almost present reasonable results.

Conclusions: Our findings show that the IDBP model is better suitable for evaluating accident datasets than other models. Motorcycle accidents, pedestrian accidents, left turn deviance, and dangerous speeding were all significant variables in the IDBP death model, and these parameters were linked to accident mortality.

Keywords: Road Traffic Accident, Death, Injury, Bivariate Regression

Backgrounds

Today, one of the biggest challenges in the world, especially in developing countries, is traffic accidents and their consequences. Road accidents include a variety of elements and conditions. Some of this data is geographically specific, while others are descriptive of accidents

[1], which are caused by driver behavior, road characteristics, and environmental variables [2–5]. Also, road accidents are one of the most serious public health issues in the world [6], with an estimated 1.2 million people killed and 50 million people wounded each year [7]. According to forensic statistics in Iran, the average traffic fatality over the past ten years has been 22,185 individuals [8]. Moreover, about 70% of accidents in this country have occurred on roads and suburban roads, and usually, the victims are healthy young people, and about 10% of them died in accidents [9]. According to United Nations data,

*Correspondence: biostat.kums@gmail.com

² Department of Biostatistics, Faculty of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran

Full list of author information is available at the end of the article



33 people have perished in traffic accidents for every 10,000 automobiles [10]. Accident-related fatalities and injuries have been recognized as a worldwide problem, and traffic safety concerns have been a significant concern since the dawn of the vehicle age, over a century ago [11]. The cost of fatalities and injuries from traffic accidents has a significant impact on society. In recent years, researchers have paid more attention to determining the factors affecting the severity of driver injuries caused by traffic accidents [12–14]. The number of traffic accidents and their effects, especially casualties, justify the importance of analyzing the factors affecting their occurrence [15–17].

Traffic control and research centers typically use descriptive statistical techniques and graphs, such as frequency tables, bar charts, and histograms to organize the number of casualties in road accidents. This statistical approach provides only a descriptive report of the rate of accidents, and based on this information, no estimate and prediction can be made. It goes without saying that having a good model for predicting and evaluating these implications is critical. It's also crucial to figure out how persons, the environment, road conditions, and vehicle types affect traffic accidents. In such analyzes, the response variable measurements are counted and the important issue is whether the counting data have a normal distribution in terms of their quantitative scale, and whether these data can be modeled using linear regression? Various studies have shown that counting variables meet the following conditions: they have a counting nature, i.e. there are no negative numbers and the amplitude of data changes starts from zero. They're also often skewed, thus they don't have a normal distribution and may be sparsely dispersed. As a result, these variables fail to fulfill the criteria for linear regression [18]. Hence, for better description and modeling, it is necessary to use the nonlinear count regression models are used, such as Poisson regression, negative binomial regression, and quasi Poisson regression [18, 19]. Although the distribution of univariate discrete data is well developed, there may be more than one response variable in some studies, such as traffic accidents analysis [20, 21]. On the other hand, road accidents have consequences such as death and injury, and the analysis of the simultaneous of these two responses has been rarely used in studies. Because the marginal distributions are not independent, it is occasionally required to utilize a bivariate distribution for correlated bivariate data.

In recent years, the multivariate model with two or more replies and counting data has gotten a lot of attention [22]. In fact, in addition to univariate models, bivariate regression models have been proposed for the

analysis of two correlated response variables. These models provide sufficient flexibility by allowing two correlated response variables that different predictors influence. Besides, a bivariate model is more useful for inference and prediction purposes because it allows us to correctly determine the dependencies between two dependent variables [23]. In recent years, different models were developed that each has some advantages and disadvantages and may present different results in different situations [24]. The bivariate count regression has shown to be challenging, because of correlation structure is not a specified parametric class of distributions.

The bivariate Poisson model is the most extensively used, although it has the same constraints as the univariate Poisson model, and the variance estimates of the model are influenced when there is over-dispersion or under-dispersion in the data. Furthermore, the presence of negative correlation and heterogeneity in variance decreases the model's efficiency [23]. Bivariate negative binomial (BNN) regression, Dirichlet negative multinomial (DNM) regression model, and diagonal inflated bivariate Poisson (DIBP) regression are introduced in the literature for the solution of these issues and have better performance to describe bivariate responses that are interdependent and also have the feature of hyper diffraction [25, 26]. Therefore, this study aimed to estimate and compare the parameters of these models to identify the factors affecting the number of mortality and the number of injured in road accidents.

Some previous studies to analyzing accident data

Over the years, a number of count regression models have been used to analyze accident data. The univariate Poisson regression and negative binomial models were utilized as a starting point for the majority of the models employed in the area. For example, in 2018, Mandacar and colleagues [27] conducted research on traffic-related deaths and severe injuries in Goiânia, Brazil. They used Poisson regression to fit the data, and then univariate analysis was used to analyze it. Moreover, recently Sagamiko and Mbare in 2021 implemented study for Modelling Road Traffic Accidents Counts in Tanzania. They used univariate Poisson regression for injuries and death accidents and ignored correlation between them [28].

In fact, accident variables, including to different levels, such as injuries and fatalities that are usually affected by big correlations and analyses based on separate models are inappropriate, and bivariate models are needed. Unfortunately, the majority of research in the area of accidents has utilized models that aren't specifically created for multivariate or bivariate models, instead

considering the correlation of variables using standard models with modifications. Table 1 shows an overview of some of the research that looked at this topic.

Data sources and study population

The accident data used in this study are related to the research area of Kermanshah province. Kermanshah province, with an area of 25,009 square kilometers and a population of 1,945,227 people, is located at latitude 34.308159 and longitude 47.05732, and the geographical position is 34.308159° north and 47.05732° east in western Iran. The Kermanshah province's Traffic Accident Command and Control Center provided these figures. The characteristics of accidents are shown in Fig. 1. The majority of accidents, as seen in these graphs, have happened on major highways. The roads from Kangavar to Qasr Shirin on the Karbala highway, as well as the route from Kamyaran to Kermanshah, were the most accident-prone.

Outcome variables

In a one-year period, 3090 accidents happened on the suburban roads of Kermanshah province, which resulted in 398 deaths and 4805 injuries. Figure 1a shows the density map of injury accidents. The highest density of injury accidents was in the axes of Mahidasht to Sarmast, Kamyaran axis, and Sarpol-e-Zahab-Qasr-e-Shirin axis. 1.36 to 1.7 people per square meter of injuries occurred in these places. Figure 1b shows the density map of fatal crashes. The highest density of fatal accidents was in Mahidasht axis, Kamyaran axis, Sarpol-e-Zahab and Qasr-e-Shirin axis, Hamil axis, and Sarab Niloufar axis. 0.28 to 0.35 people per square meter of fatal accidents occurred in these places.

Predictors

Population-level variables that were included in the regression models were accident information, including:

Date, time and place of accident, longitude, and latitude of accident place, the crash location latitude and longitude coordinates, obtained by GPS or mapping type of accident (fatal and injury), number of fatalities, number of injuries, accident mode (multiple vehicles (Collisions between motor vehicles), off-road (Off-road motor vehicle), fixed object (Collisions between a motor vehicle and fixed object), pedestrian (Collisions between a motor vehicle and pedestrian), Collisions between a motor vehicle and motorcycle, Overturning (Overturning a motor vehicle), etc.), the cause of the accident (deviation to the left, speeding (high speed motor vehicle), reversing (Vehicle rotation), fatigue and drowsiness (Drivers who don't get enough sleep), lack of attention of driver to the front, non-compliance with the right of way, non-compliance with the longitudinal and transverse distance, technical defects, etc.), and the type of vehicle (freight, personal ride, public passenger and motorcycle).

Software

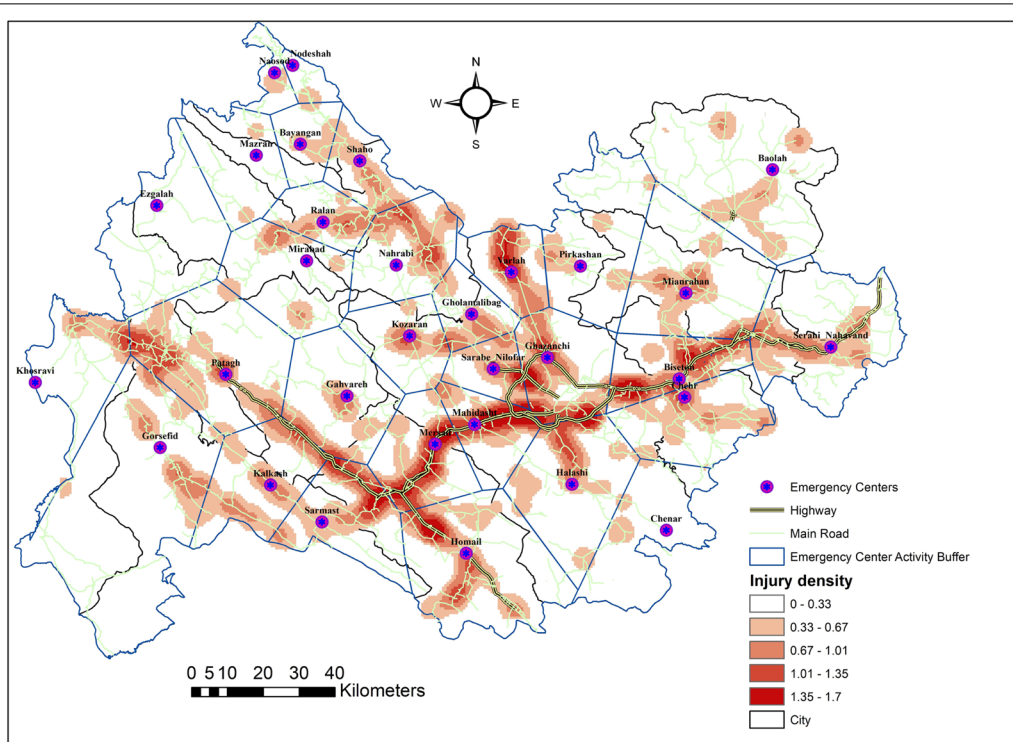
Data were analyzed using ArcGIS 10.2 and R4.0.5 software. The R packages to be used included ggplot2, msme, Count, moments, Mass, MGLM, BPglm and bivpois. Significant level in this study is considered at 0.05.

Statistical analysis

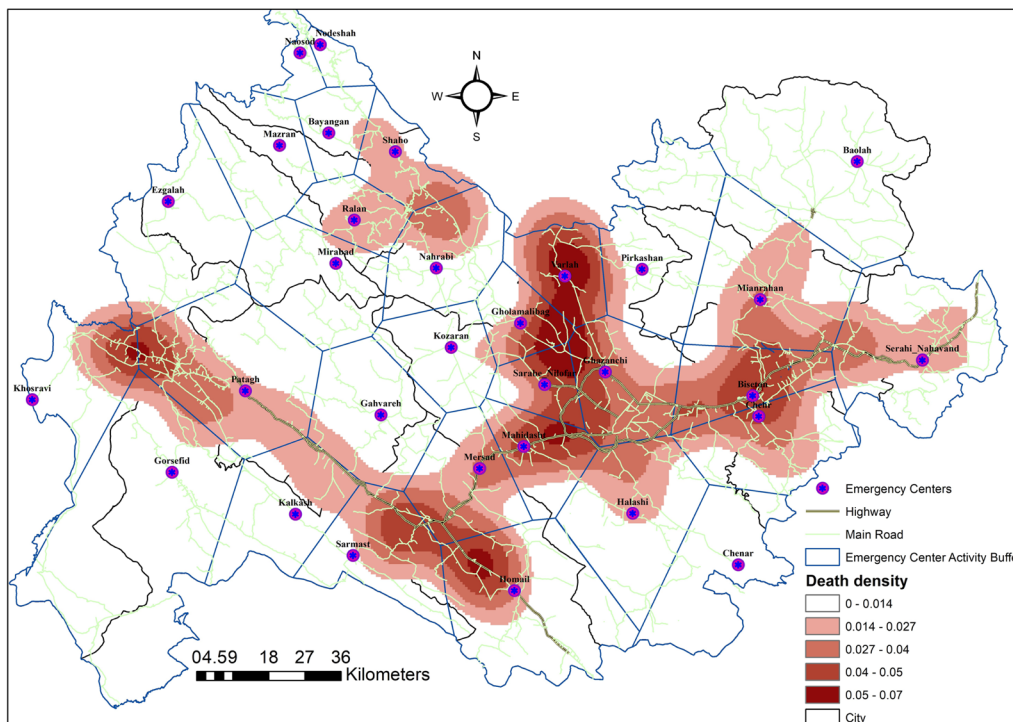
In recent years, joint modeling of two or more counting outcomes has gotten a lot of interest. When two counting variables are associated and need to be estimated

Table 1 Summary of some previous studies to analyzing accident data

Model type	Year	Conclusion
A joint model with Weighted risk score to combine crash count and crash severity [29]	2020	using of crash severity and crash count amended the accuracy of prediction model
A bivariate Bayesian hierarchical extreme value model for traffic conflict-based crash estimation [30]	2020	The bivariate model estimate regression coefficients more precisely than univariate models
Bayesian multivariate hierarchical spatial joint model [31]	2018	This model has a better fit for the crash data compared to the univariate alternative model
Copula-Based Joint Model of Injury Severity and Vehicle Damage in Two-Vehicle Crashes [32]	2015	On the basis of goodness-of-fit statistics, the Gaussian copula model that was calculated interrelationships between injury severity and vehicle damage was suitable
using the random-parameters tobit model for factors affecting highway accident rates [33]	2012	The empirical results show that this model was proper fit to the data
A joint-probability approach to crash prediction models [34]	2011	Joint probability model that modeled Crash occurrence and severity simultaneously, shown the good fit for data
Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity [35]	2007	The results show multivariate model that accounted correlation of variables, was achieved more accurate estimates



a: Density map of injuries accidents in march2020 to march2021in Kermanshah



b: Density map of fatalities accidents in march2020 to march2021in Kermanshah

Fig. 1 a Density map of injuries accidents in march2020 to march2021in Kermanshah. b Density map of fatalities accidents in march2020 to march2021in Kermanshah

together, bivariate counting models may be utilized. Various models have been devised, each with its own set of benefits and drawbacks that may provide different outcomes in different scenarios. We have compared univariate Poisson, negative binomial regression and the four bivariate count regression models for accident data used commonly, which summary of them are mentioned below and reader can refer to their references for more details.

Poisson regression

In statistics and probability, the Poisson process is used to represent a sequence of events with discrete values under the following conditions, and the random variable in this process has a Poisson distribution.

- 1) The number of event occurs in a specific time or place interval.
- 2) The probability of each event occurring is independent of the occurrence of other events, that is, the events are independent of each other.
- 3) The average number of events per time period is constant.
- 4) It is not possible for two events to occur at the same time.

A discrete random variable X is said to be a Poisson random variable with parameter λ is shown as $X \sim p(\lambda)$ that

$$p(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}; x = 0, 1, \dots, \infty, \lambda > 0$$

$$E(x) = \text{Var}(x) = \lambda$$

Almost all numerical models have the basic structure of a linear model, and in Poisson regression, the difference is that the equation on the left is expressed as a logarithm:

$$\ln(\lambda) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

Note that there is no linear relationship between the model and predictive (independent) variables such as the linear model, and that there is a linear relationship between the natural logarithm of λ and the predictor variables. An important feature of the natural logarithm link function for counting data is that it ensures that predicted values will always be positive. A process called maximum likelihood estimation (MLE) may be used to estimate β values.

Bivariate poisson regression

The bivariate Poisson regression models are the most widely used for bivariate counts data. These models [36] define the correlation structure through the trivariate reduction method, where the pair of dependent variables is specified using three random variables. Consider a pair of random variables (X, Y) that have a common distribution as follows:

$$\begin{aligned} P(X = 0, Y = 0) &= p_{00}, P(X = 1, Y = 0) \\ &= p_{10}, P(X = 0, Y = 1) \\ &= p_{01}, P(X = 1, Y = 1) = p_{11} \end{aligned}$$

$$\text{that } p_{00} + p_{10} + p_{01} + p_{11} = 1$$

First, bivariate Poisson distribution is constructed that the probabilities are denoted as $p_{11} = \frac{\lambda_{11}}{n}, p_{01} = \frac{\lambda_{01}}{n}, p_{10} = \frac{\lambda_{10}}{n}$ then joint distribution for n independent vectors $(X_1, Y_1), \dots, (X_n, Y_n)$ is defined $P(\sum_{i=1}^n X_i = k, \sum_{i=1}^n Y_i = l) = \sum_{\delta=\max(k+l-n,0)}^{\min(k,l)} \frac{n!}{(n-(k+l)+\delta)(k-\delta)!(l-\delta)!} = \left(1 - \frac{\lambda_{10} + \lambda_{01} + \lambda_{11}}{n}\right)^{n-(k+l)+\delta} \left(\frac{\lambda_{10}}{n}\right)^{k-\delta} \left(\frac{\lambda_{01}}{n}\right)^{l-\delta} \left(\frac{\lambda_{11}}{n}\right)^\delta$

In the above equation, the right side converges to the following equation:

$$\frac{\lambda_{10}^{k-\delta} \lambda_{01}^{l-\delta} \lambda_{11}^\delta}{(k-\delta)!(l-\delta)!\delta!} e^{-(\lambda_{10} + \lambda_{01} + \lambda_{11})}$$

Then the partial distribution of the sum vector (X, Y) is as follows which gives the marginal distribution of BP distribution:

$$P(X = k, Y = l) = \sum_{\delta=0}^{\min(k,l)} \frac{\lambda_{10}^{k-\delta} \lambda_{01}^{l-\delta} \lambda_{11}^\delta}{(k-\delta)!(l-\delta)!\delta!} e^{-(\lambda_{10} + \lambda_{01} + \lambda_{11})}$$

If $(X_i, Y_i) \sim \text{BP}(\lambda_{1i}, \lambda_{2i}, \lambda_{3i})$ then BP regression model is defined as following that w_{ki} is the vector of i th predictor and β_k is the vector of k th regression coefficient.

$$\log(\lambda_{1i}) = w_{1i} \top \beta_1,$$

$$\log(\lambda_{2i}) = w_{2i} \top \beta_2,$$

$$\log(\lambda_{3i}) = w_{3i} \top \beta_3$$

Bivariate negative binomial regression

Based on a similar correlation structure in the BP regression model, Famoye [37] introduced a BNB regression model to analyze bivariate data, addressing over-dispersion in it. This model can be used to have a negative, zero, or positive correlation. It uses separate dispersion parameters for each marginal distribution, which can be shown as follows:

$$p(X, Y) = \left\{ \prod_{k=1}^2 \binom{m_k^{-1} + y_k - 1}{y_k} \left(\frac{\mu_k}{m_k^{-1} + \mu_k} \right)^{y_k} \left(\frac{m_k^{-1}}{m_k^{-1} + \mu_k} \right)^{m_k^{-1}} \right\} \times p(X, Y)$$

$$= \left\{ \prod_{k=1}^2 \binom{m_k^{-1} + y_k - 1}{y_k} \left(\frac{\mu_k}{m_k^{-1} + \mu_k} \right)^{y_k} \left(\frac{m_k^{-1}}{m_k^{-1} + \mu_k} \right)^{m_k^{-1}} \right\}$$

$$\times \{1 + \lambda(e^{-y_1} - c_1)(e^{-y_2} - c_2)\}$$

$$c_k = \{(1 - \theta_k)/(1 - \theta_k e^{-1})\}^{m_k^{-1}} \theta_k = \mu_k/(m_k^{-1} + \mu_k)$$

$$k = 1, 2$$

That m_k is the dispersion parameter for the negative binomial distribution.

Dirichlet negative multinomial regression

DNM distribution [38] is a discrete multivariate distribution that supports the count variables. Suppose (Y_1, \dots, Y_m) are independent Poisson variables with mean vector $(\lambda_1 b, \lambda_2 b, \dots, \lambda_m b)$. If gamma (Y_0, Y_0) that Y_0 is a shape parameter. If λ_i are fixed and known, then (Y_1, \dots, Y_m) has a NM distribution and if λ_i is positive random variable, it has the DNM distribution. Mass function for DNM distribution with parameters $Y_0, \lambda_1, \dots, \lambda_m$ is

$$\frac{\Gamma(Y_*)}{\Gamma(Y_0) \prod_{j=1}^m \Gamma(Y_{j+1})} \times \frac{\Gamma(\lambda_g)}{\prod_{j=1}^m \lambda_j} \times \frac{\prod_{j=0}^m \Gamma(Y_j + \lambda_j)}{\Gamma(Y_* + \lambda_*)}$$

$$Y_* = Y_0 + \dots + Y_m$$

$$\lambda_* = \lambda_1 + \dots + \lambda_3$$

Diagonal inflated bivariate poisson regression

The DIBP regression [27] is modified of BP regression that addressed over-dispersion and under-dispersion.

Under the PB regression model, the DIBP model is defined as follows:

$$f_{IBP}(X, Y) = \begin{cases} (1 - p)f_{BP}(X, Y|\lambda_1, \lambda_2, \lambda_3) & X \neq Y \\ (1 - p)f_{BP}(X, Y|\lambda_1, \lambda_2, \lambda_3) + pf_D(X|D) & X = Y \end{cases}$$

$f_D(X|D)$ is discrete probability distribution that proper choice for it, is the Poisson or geometric distribution.

Comparing regression models

Mean Square Error (MSE) is a measure of the error model by average squared differences between the estimated response value from the fitted model and observed value. If the fitted model is acceptable, the observed and their estimated values will be close and MSE to be small. Besides, Akaike Information Criterion (AIC) index that measures the basis of log-likelihood is used for comparing model and smaller values to be well.

Overdispersion

OV is one of the most important features of counting data. To investigate the OV in univariate analysis, z test was used in which it is assumed that the variance of a variable is $var(Y) = \mu + c \times f(\mu)$ that if c is zero, there is no OV. Furthermore, for bivariate analyzes, OV will be evaluated based on the chi-square test [38, 39].

Results

Frequency of injuries and death

Based on density characteristics, accidents were split into 125 regions. The distribution of the number of injuries and fatal accidents was observed using a histogram plot (Fig. 2). The data is not symmetrically distributed and is biased to the right, as can be observed. The number of wounded in each location is 38.41(60.034) individuals on average (SD), with a median of 18.5 people. Moreover, the mean (SD) for the number of fatal in each area is 3.5(5.91) people with a median of 1 person. The correlation between the number of injuries and deaths in

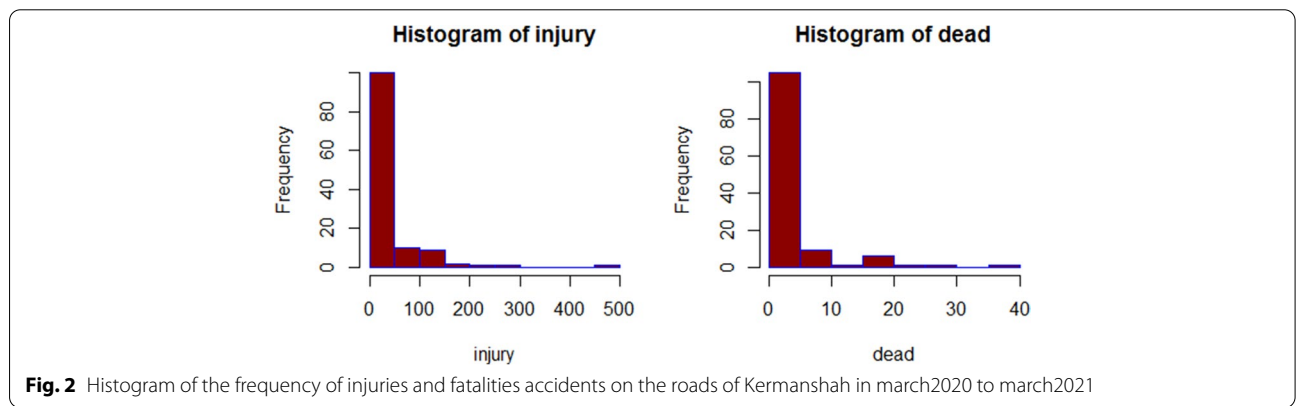


Fig. 2 Histogram of the frequency of injuries and fatalities accidents on the roads of Kermanshah in march2020 to march2021

terms of road accidents was 0.856 and shows that these two response variables positively correlate with each other, shown in Fig. 3. This research looked at seventeen factors, including the frequency of accidents, accident states, and accident status. Overturning and collisions of cars with each other or multiple vehicles resulted in the most accidents, whereas collisions with a stationary object resulted in the fewest. Most of the vehicles collided were happened in Kamyaran route, which also has the highest number of injuries. The highest number of deaths is related to the accident of several vehicles, the most dangerous type of which is a face-to-face collision on non-separate two-way axes, antenna intersections, and non-standard U-turns.

Evaluate overdispersion

As shown in Table 2, the OV test was significant for both response variables (death and injury) in univariate analysis and injury variables in bivariate analysis.

There was significant OV in death and injury univariate model and in bivariate death model.

Fitted count regression models

Univariate count models

First, results consider and compare two common univariate models (P regression and NB regression) for the number of injuries and deaths.

P regression and NB regression were used to investigate the relationship between the number of injured with type and cause of the accident. As can be seen in Table 3, the results of NB regression and P regression are different due to the overdispersion feature of the data and therefore the NB regression gives more reliable results. The results show that the multi-vehicle collision, riding with riding, overturning, and pedestrian and riding a motorcycle had a significant relationship with the number of injured in the accident. P regression and NB regression

Table 2 Test for over-dispersion in univariate and bivariate regression models

Outcomes	Univariate			Bivariate	
	Z	alpha	p-value	Chi-square	p-value
death	2.64	0.746	0.004	59.649	0.0078
injured	2.013	0.632	0.001	13.284	0.9998

were also used to investigate the relationship between the number of death and the type and shape of the accident. In regression models, multi-vehicle collision, ride-on-ride, pedestrian and motor-ride had a significant relationship with the number of death.

Bivariate count models

This paper considers and compares four different candidate bivariate models (BP regression, BNB regression, DIBP regression, and DNM models) for the number of injuries and deaths. Results indicated that the most of predictors are significant under Poisson regression model for injury however some of predictors was found significant using bivariate count regression model so the results of univariate and bivariate models are different. Deviance to the left and colliding with a pedestrian were two factors that became significant in both death and injury models (Table 4).

Comparison of models

In bivariate models, except for the DNM model, there is a reasonable decrease in the AIC measures of the saturated model compared to the reduced model for the other three models. Moreover, the results of MSE suggest that the DIBP model is a better fit for these data. As shown in Table 5 for the injuries model, MSE is lowest, respectively

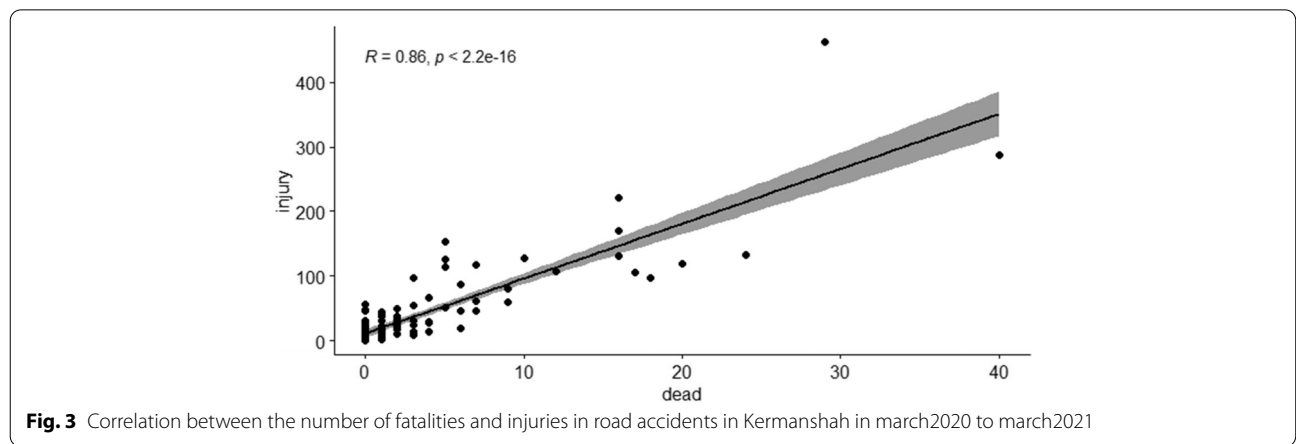


Fig. 3 Correlation between the number of fatalities and injuries in road accidents in Kermanshah in march2020 to march2021

Table 3 Univariate regression count model: parameter estimation for injuries and deaths

Factor	Univariate Regression			
	P		NB	
	Death	Injured	Death	Injured
Accident number	0.0485 (0.0451)	-0.0696 (0.0134) ^a	0.0485 (0.0451)	-0.0792 (0.0447)
Fixed objects	-0.0172 (0.0857)	0.1838 (0.0262) ^a	-0.0172 (0.0857)	0.2132 (0.0813) ^a
Multiple-vehicle collision	0.1623 (0.0965)	0.3745 (0.0290) ^a	0.1623 (0.0965)	0.3575 (0.0928) ^a
Head-On Collision	-0.0565 (0.0606)	0.0978 (0.0183) ^a	-0.0565 (0.0606)	0.0348 (0.0678)
Truck collision	-0.0605 (0.0734)	0.0699 (0.0221) ^a	-0.0605 (0.0733)	0.1241 (0.0697)
vehicle collision	-0.0194 (0.0525)	0.1096 (0.0165) ^a	-0.0194 (0.0525)	0.1221 (0.0520) ^a
Motorcycle collision	-0.0264 (0.0582)	0.1614 (0.0177) ^a	-0.0264 (0.0582)	0.1747 (0.0569) ^a
Pedestrian	-0.1912 (0.0705) ^a	-0.0034 (0.0199)	-0.1912 (0.0705) ^a	0.0461 (0.0623)
Vehicle Rollover	-0.0424 (0.0445)	0.1217 (0.0128) ^a	-0.0424 (0.0445)	0.1293 (0.0435) ^a
Other type of accidents	0.0195 (0.0620)	0.1063 (0.0173) ^a	0.0195 (0.0620)	0.1194 (0.0620)
turn left without using the turn signal	0.0413 (0.0244)	-0.0104 (0.0077)	0.0414 (0.0245)	-0.0122 (0.0291)
Over-speeding	-0.0081 (0.0317)	-0.0387 (0.0091) ^a	-0.0816 (0.0317)	-0.0152 (0.0272)
Passing without signaling	-0.0495 (0.0787)	-0.0652 (0.0249) ^a	-0.0496 (0.0787)	-0.1660 (0.0843) ^a
Lack of attention to the front	0.1044 (0.0259) ^a	0.0195 (0.0081) ^a	0.1044 (0.0259) ^a	-0.0092 (0.0266)
neglect of the right of priority	0.1589 (0.0498) ^a	0.0195 (0.049)	0.1588 (0.0498) ^a	0.0332 (0.0491)
transverse distance	-0.1391 (0.1275)	-0.1055 (0.0367) ^a	-0.1391 (0.1275)	-0.1807 (0.113)
Breaking traffic rules	-0.1966 (0.0611) ^a	-0.1417 (0.017) ^a	-0.1967 (0.0612) ^a	-0.1058 (0.0593)

^a Significant at 0.05 level

for DIBP (137.87), BNB (289.46), BP (412.36) and DNM (3640.89) models. These results are established for dead models.

Discussion

Forecasting based on facts, accessible parameters, and available information are the major aims of modeling and categorization in statistics, and there are numerous statistical approaches for achieving these goals and modeling. Because of the rising number of road accidents in the nation, having an indicator that displays the current state of road safety may be required for controlling road traffic accidents, thus study in this area is critical. Most fatalities in traffic accidents are caused by dangerous drivers and people living in low- and middle-income

countries, where transportation is increasing. Many reasons, such as rapid urbanization, poor road safety, poor law enforcement, distracted or tired driving, drug or alcohol use, speeding, and not wearing seat belts and helmets, contribute to traffic accidents [40]. This data was shown that 45% of accidents were due to speeding, 23.33% due to left turn and 10.83% due to neglect of the right of priority. Also, the condition of the roads also affects the occurrence of traffic accidents. About 50% of the dead and 37% of the injured in the province occurred in accidents of 9 main routes, of which five routes are non-separate and narrow. Two axes are widening whenever their completion is delayed, it is directly effective in increasing the number of deaths and injuries in the province. The two alternative routes are national and

Table 4 bivariate count regression modes:l parameter estimation for injuries and deaths

Factor	Bivariate Regression								
	BP		BNB		IDBP		DNM		Wald
	Death	Injured	Death	Injured	Death	Injured	Death	Injured	
Accident number	0.089 (0.048)	0.1195 (0.0129) ^a	0.2327 (0.1825)	1.0879 (0.5946)	0.4777 (0.1179)	-0.1385 (0.4224)	0.0468	0.3986	2.653
Fixed objects	0.0535 (0.0916)	-0.1666 (0.0259) ^a	-0.4258 (0.3297)	-0.2979 (1.0742)	0.2615 (0.1168)	0.3861 (0.2872)	0.6478	0.5948	0.9878
Multiple-vehicle collision	0.2588 (0.1030) ^a	0.0433 (0.0273)	0.2596 (0.3807)	-1.1074 (1.2403)	0.0202 (0.1567)	0.1491 (0.1116)	0.6759	0.8604	2.2006
Head-On Collision	0.0886 (0.0637)	0.1677 (0.0173) ^a	0.2323 (0.2789)	0.2515 (0.9089)	-0.7134 (0.3658)	0.6766 (0.1839)	-0.7321	0.8818	3.9526
Truck collision	-0.0881 (0.0796)	-0.1365 (0.0224) ^a	0.0076 (0.2854)	1.6612 (0.9298)	-0.3457 (0.2889)	0.2850 (0.0456)	0.0609	1.0563	3.4932
vehicle collision	0.0619 (0.0553)	0.1577 (0.0158) ^a	-0.0567 (0.2116)	1.9764 (0.6894) ^a	-0.3288 (0.1263)	0.3495 (0.3348) ^a	-0.5637	0.4786	2.8534 ^a
Motorcycle collision	0.035 (0.0617)	0.1799 (0.0172) ^a	0.3681 (0.2309)	0.5088 (0.7525)	0.8051 (0.4824) ^a	0.3131 (0.5022)	0.3233	0.1914	2.9601
Pedestrian	0.256 (0.0763) ^a	0.2041 (0.0197) ^a	0.6339 (0.2518) ^a	1.6878 (0.8206)	-0.8117 (0.0716)	-0.7549 (0.369)	0.7327	0.6152	3.4130 ^a
Vehicle Rollover	0.0629 (0.0474)	0.1335 (0.0116) ^a	0.2855 (0.1775)	0.8541 (0.5784)	0.6751 (0.0545)	0.386 (0.4372)	0.4288	0.3233	4.5605
Other type of accidents	0.0039 (0.0656)	0.0526 (0.0165) ^a	0.085 (0.2551)	0.5196 (0.8312)	-0.6574 (0.2267)	0.6502 (0.4032)	0.0471	0.0781	0.1137
turn left without using the turn signal	0.0476 (0.0255)	0.0796 (0.0077) ^a	0.3664 (0.1212) ^a	1.2507 (0.3948) ^a	0.8269 (0.2394) ^a	0.7106 (0.2114) ^a	0.1975	0.1424	0.3086 ^a
Over-speeding	0.0409 (0.0336)	0.0247 (0.0093) ^a	0.1165 (0.1098)	0.2509 (0.358)	0.9842 (0.0294)	0.3647 (0.1418)	0.2769	0.3049	1.8007
Passing without signaling	-0.8884 (0.0842)	0.0067 (0.0240)	-0.0233 (0.358)	1.1216 (1.1664)	-0.5923 (0.1172)	0.6367 (0.2496)	0.1502	0.2328	0.3381
Lack of attention to the front	0.1324 (0.0272) ^a	0.099 (0.0077) ^a	0.3196 (0.1092) ^a	0.6132 (0.3559)	-0.1372 (0.0598)	0.7405 (0.1345)	0.2407	0.1499	7.7306
neglect of the right of priority	-0.1869 (0.0530)	-0.1832 (0.0148) ^a	0.6628 (0.1991) ^a	0.9063 (0.6489)	-0.0981 (0.0522)	0.9309 (0.1579) ^a	0.1845	0.3203	3.3667 ^a
transverse distance	-0.0768 (0.135)	0.1077 (0.0344)	0.4208 (0.458)	-0.3148 (1.4921)	0.5975 (0.1079)	0.1234 (0.1693)	0.0854	1.0061	1.3972
Breaking traffic rules	-0.2212 (0.0641) ^a	0.0623 (0.0174) ^a	0.2463 (0.2395)	1.8660 (0.7805)	0.5125 (0.4309)	0.3545 (0.0066)	0.3188	0.4517	2.4747

^a Significant at 0.05 level

Table 5 Goodness of fit test statistics for univariate and bivariate count regression models

Index		Univariate Regression			Bivariate Regression		
		P	NB	BP	BNB	IDBP	DNM
MSE	Injury death	0.258	0.267	412.36	289.46	137.87	3640.89
		0.261	0.261	12.16	4.012	1.71	36.62
AIC	Fitted model	411.59 ^a	413.59 ^a	1402.48	963.2	1625	397.45
	Reduced model	603.63 ^a	2768.1 ^a	1897.54	986.94	2689	398.34

^a For death model

international highways, which must protect the remaining 15 kilometers between Kermanshah and Islamabad, as well as expand the number of roadside resorts and

electronic control systems. The data showed that 45% of accidents were due to speeding, 23.33% due to left turn, 10.83% due to non-compliance with the priority.

Identifying the factors affecting deaths and injuries from road accidents is essential in health system policy-making in reducing mortality. It is necessary to use statistical analysis, such as the count nonlinear regression model to better describe and analyze the number of accident data and find the impact of humans, road conditions, and vehicle type on traffic accidents. On the other hand, road accidents have consequences such as death and injury, and the simultaneous analysis and description of these two responses have rarely been used in studies. Therefore, the purpose of this study is to compare the fit of several bivariate regression models applied to road accident data and compare them. BP model is the most widely used model for two-variable counting. However, as demonstrated in the findings, the univariate Poisson model's limitations apply here as well, and model variance estimations are influenced by over- or under-dispersion in the data. Furthermore, the presence of negative correlation and heterogeneity in variance diminishes the model's effectiveness. The BNB regression model is used to represent paired counting data with interdependent response variables and the OV feature [18]. But, the issue of negative correlation still applies here. Therefore, a major drawback of the above models is their ability to model data only with a positive correlation. Besides, because they are Poisson marginal distributions, they cannot model over-scattering or low-scattering. In the case of sparse count data, Poisson mixed models are potentially useful, and some other models make negative correlations possible. However, such models involve difficult and complex calculations to estimate. Another model is the IDBP model, which is computationally feasible and allows for OV and negative correlation. The DNM model is a good marginal regression model for counting data that takes into account differences between and within units. As a result, if a marginal model is sought, the DNM model offers an appropriate structure. In particular, having separate mean parameters for each component and two variance parameters makes this model suitable with unbalanced panel counting data with a stable covariance structure.

Our results show the suitability of the IDBP model for analyzing accident datasets. In the IDBP death model, the variables of a motorcycle accident, pedestrian accident, left turn deviance, and unsafe speeding was significant, and these factors were related to accident mortality. In IDBP injury model, the variables of a driving accident, pedestrian accident, left turn deviance, and non-observance of the right of priority was significant, and these factors were related to injury in the accident. Major pedestrian accidents are in terms of the compulsion of pedestrians in traffic from the environment or a place

originally designed for using vehicles. Unfortunately, the high number of pedestrian accidents in the country and the number of casualties and disabilities caused by it have created many obvious and hidden problems for society. Overtaking and veering to the left are the province's most dangerous driving violations. Overtaking violations and left-hand deviations occur on non-separate two-way roadways that result in head-on collisions. For this reason, seven axes of the province are vital, and as long as their return routes are not separated, this sort of violation and the consequent losses will occur.

The findings of the present study are consistent with the results of Al-Ghamdi's study of traffic accident data. In this study, logistic regression was used to examine the contribution of several variables in the severity of the accident. Out of 9 independent variables obtained from police accident reports, two variables; the location and cause of the accident were significantly related to the severity of the accident [41]. Abdissa Aga et.al used univariate count regression models to analysis the factors associated with the number of human deaths from road traffic accidents. The study aimed to identify the potential factors associated with the number of human deaths by road traffic accidents in the Oromia Regional State, Ethiopia. The hurdle Poisson regression model was shown most appropriate model from other common count models. The results are shown that the number of deaths due to driving in an illegal way compared to drivers denying priority to pedestrians was lower [42].

Unfortunately the number of studies that compared bivariate count models is very small. The results obtained from comparison of models in this study are consistent the study of Famoye in 2010 which showed that BNB regression model performs better than the BP regression model [43]. Rafiqul et.al, introduced a bivariate zero-truncated Poisson regression model based on a conditional model for accident data. Two correlated outcome variables were frequency of cars and number of casualties in accidents. Based on goodness of fit index , AIC, BIC and deviance, results are shown that the proposed full model provides the best fit [44].

Conclusions

The data show that amending in legislation, vehicle standards and access to post-accident care has been developed. However, this progress has not been rapid enough to offset the consequences of traffic accidents that occur in many parts of the world specially developing country [40].

The objective of this study was to evaluate the statistical model of road accidents in Iran. The P and NB regression

model was used for univariate fitting the relationship between road accidents injuries and fatalities and their contributing factors which are reckless driving, careless pedestrians, high speed, defective motor vehicles, motor cyclists and other factors including slippery road and poor visibility. The OV test was carried out and it was shown, there was over-dispersion in the data and so NB regression was more reliable. However because there were 2 correlated count responses (injury and death), bivariate count models was used to improve results. Finally, basis on GF test and MSE index, IDBP model is a better fit for this data and vehicle collision, turn left without using signal and neglect of the right of priority for injured response and motorcycle collision and turn left without using signal for dead response are significant.

Abbreviations

P: Poisson; NB: Negative Binomial; BP: Bivariate poisson; BNN: Bivariate negative binomial; DNM: Dirichlet negative multinomial; DIBP: Diagonal inflated bivariate Poisson; MSE: Mean Square Error; AIC: Akaike Information Criterion; OV: Overdispersion.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12873-022-00686-6>.

Additional file 1.

Additional file 2.

Additional file 3.

Acknowledgements

The authors would like to express their appreciation towards the financial support of the Research and Technology Department of Kermanshah University of Medical Science This study was approved by (reference number: IR.KUMS.REC.1400.560) and also are grateful to the reviewers that their comments improved the quality of the paper.

Authors' contributions

Shahsavari S, Mohammadi A and Zeini F performed the analysis and Shahsavari S and Tabatabaei SM interpreted the results. Shahsavari M, Mostafaei S, Zhaleh M and Zereshki E drafted the paper. All authors read and approved the final manuscript.

Funding

No any sources of funding for the our research work and their role in the design of the study and collection, analysis, interpretation of data, and in writing the manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Health Information Technology, Faculty of Allied Sciences, Kermanshah University of Medical Sciences, Kermanshah, Iran. ²Department of Biostatistics, Faculty of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran. ³Inflammation Research Center, Tehran University of Medical Sciences, Tehran, Iran. ⁴Department of Medical Informatics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. ⁵Department of Anatomy and Cell Biology, Medicine Faculty, Kermanshah University of Medical Sciences, Kermanshah, Iran. ⁶Imam Ali Hospital Heart Center, Kermanshah University of Medical Sciences, Kermanshah, Iran.

Received: 27 April 2022 Accepted: 1 July 2022

Published online: 18 July 2022

References

- Nantulya VM, Reich MR. The neglected epidemic: road traffic injuries in developing countries. *BMJ*. 2002;324(7346):1139–41.
- Onate-Vega D, Oviedo-Trespalacios O, King MJ. How drivers adapt their behaviour to changes in task complexity: the role of secondary task demands and road environment factors. *Transport Res F: Traffic Psychol Behav*. 2020;1(71):145–56.
- Wu J, Xu H. Driver behavior analysis on rural 2-lane, 2-way highways using SHRP 2 NDS data. *Traffic Inj Prev*. 2018;19(8):838–43.
- Ellison AB, Greaves S, Bliemer M. Examining heterogeneity of driver behavior with temporal and spatial factors. *Transp Res Rec*. 2013;2386(1):158–67.
- Pakgohar A, Tabrizi RS, Khalili M, Esmaeili A. The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach. *Procedia Computer Science*. 2011;1(3):764–9.
- Kulharni J. Public Health Issue Related to Road Traffic Crashes (RTCs). *Int J Collab Res Int Med Public Health*. 2021;13(2):1–6.
- World Health Organization. Global status report on road safety 2015. World Health Organization; 2015. <https://apps.who.int/iris/handle/10665/189242>.
- Esmaeili A, Farashi B. A strategic model for developing the culture of traffic. *DANESH-E-ENTEZAMI*. 2008;10(39):104–27.
- Pakgouhar AR, Khalili M, Safarzadeh M. The consideration of human factor's role in occurrence and aggravation of road accidents based on the regression models Lr and cart. *Traffic Management Studies*. 2009;4(13):49–66.
- Ruikar M. National statistics of road traffic accidents in India. *J Orthop Traumatol Rehabil*. 2013;6(1):1.
- Elmansouri O, Almhroog A, Badi I. Urban transportation in Libya: an overview. *Transport Res Interdiscip Perspect*. 2020;1(8):100161.
- Mohammed AA, Ambak K, Mosa AM, Syamsunur D. A review of traffic accidents and related practices worldwide. *The Open Transportation Journal*. 2019;13(1).
- Wijnen W, Weijermars W, Schoeters A, van den Berghe W, Bauer R, Carnis L, Elvik R, Martensen H. An analysis of official road crash cost estimates in European countries. *Saf Sci*. 2019;1(113):318–27.
- Byrne JP, Mann NC, Dai M, Mason SA, Karanicolas P, Rizoli S, Nathens AB. Association between emergency medical service response time and motor vehicle crash mortality in the United States. *JAMA Surg*. 2019;154(4):286–93.
- Mannering FL, Shankar V, Bhat CR. Unobserved heterogeneity and the statistical analysis of highway accident data. *Anal Methods Accid Res*. 2016;1(11):1–6.
- Eboli L, Forciniti C, Mazzulla G. Factors influencing accident severity: an analysis by road accident type. *Transport Res Procedia*. 2020;1(47):449–56.
- De Oña J, Mujalli RO, Calvo FJ. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. *Accid Anal Prev*. 2011;43(1):402–11.
- Zhang Y, Zhou H, Zhou J, Sun W. Regression models for multivariate count data. *J Comput Graph Stat*. 2017;26(1):1–3.

19. Ajiferuke I, Famoye F. Modelling count response variables in informetric studies: Comparison among count, linear, and lognormal regression models. *J Informet.* 2015;9(3):499–513.
20. Famoye F. Comparisons of some bivariate regression models. *J Stat Comput Simul.* 2012;82(7):937–49.
21. Gurmu S, Elder J. A simple bivariate count data regression model. *Economics Bulletin.* 2007;3(11):1.
22. Bermudez L, Karlis D. A finite mixture of bivariate Poisson regression models with an application to insurance ratemaking. *Comput Stat Data Anal.* 2012;56:3988–99.
23. AlMuhayfith FE, Alzaid AA, Omair MA. On bivariate Poisson regression models. *J King Saud Univ-Sci.* 2016;28(2):178–89.
24. Holgate P. Estimation for the bivariate Poisson distribution. *Biometrika.* 1964;51:241–5.
25. Johnson NL, Kotz S, Balakrishnan N. *Discrete Multivariate Distributions.* New York: John Wiley & Sons; 1997.
26. Kocherlakota S, Kocherlakota K. Regression in the bivariate Poisson distribution. *Commun Statist Theory Method.* 2001;30:815–25.
27. Mandacarú PM, Rabelo IV, Silva MA, Tobias GC, MoraisNeto OL. Deaths and serious injuries due to traffic accidents in Goiânia, Brazil-2013: the magnitude and associated factors. *Epidemiologia e Serviços de Saúde.* 2018;7:27.
28. Sagamiko T, Mbare N. Modelling Road Traffic Accidents Counts in Tanzania: A Poisson Regression Approach. *Tanzania J Sci.* 2021;47(1):308–14.
29. Afghari AP, Haque MM, Washington S. Applying a joint model of crash count and crash severity to identify road segments with high risk of fatal and serious injury crashes. *Accid Anal Prev.* 2020;1(144):105615.
30. Zheng L, Sayed T. A bivariate Bayesian hierarchical extreme value model for traffic conflict-based crash estimation. *Anal Methods Accid Res.* 2020;1(25):100111.
31. Alarifi SA, Abdel-Aty M, Lee J. A Bayesian multivariate hierarchical spatial joint model for predicting crash counts by crash type at intersections and segments along corridors. *Accid Anal Prev.* 2018;1(119):263–73.
32. Wang K, Yasmin S, Konduri KC, Eluru N, Ivan JN. Copula-based joint model of injury severity and vehicle damage in two-vehicle crashes. *Transp Res Rec.* 2015;2514(1):158–66.
33. Anastasopoulos PC, Mannering FL, Shankar VN, Haddock JE. A study of factors affecting highway accident rates using the random-parameters tobit model. *Accid Anal Prev.* 2012;1(45):628–33.
34. Pei X, Wong SC, Sze NN. A joint-probability approach to crash prediction models. *Accid Anal Prev.* 2011;43(3):1160–6.
35. Park ES, Lord D. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. *Transp Res Rec.* 2007;2019(1):1–6.
36. Karlis D, Ntzoufras I. Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *J Stat Softw.* 2005;5(14):1–36.
37. Famoye F. On the bivariate negative binomial regression model. *J Appl Stat.* 2010;37(6):969–81.
38. Farewell DM, Farewell VT. Dirichlet negative multinomial regression for overdispersed correlated count data. *Biostatistics.* 2013;14(2):395–404.
39. Jung BC, Jhun M, Han SM. Score tests for overdispersion in the bivariate negative binomial models. *J Stat Comput Simul.* 2009;79(1):11–24.
40. World Health Organization. *Global status report on road safety 2018* (2018). Geneva: World Health Organization; 2019.
41. Al-Ghamdi AS. Using logistic regression to estimate the influence of accident factors on accident severity. *Accid Anal Prev.* 2002;34(6):729–41.
42. Aga MA, Woldeamanuel BT, Tadesse M. Statistical modeling of numbers of human deaths per road traffic accident in the Oromia region, Ethiopia. *PLoS ONE.* 2021;16(5):e0251492.
43. Famoye F. On the bivariate negative binomial regression model. *J Appl Stat.* 2010;37(6):969–81.
44. Chowdhury RI, Islam MA. Zero truncated bivariate Poisson model: Marginal-conditional modeling approach with an application to traffic accident data. *Appl Math.* 2016;7(14):1589.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

