Poisson Regression Model of frequency and severity of Traffic collisions in rural roads

M. Asadamraji[1*] S. Seyedabrishami[2]

M. Saffarzadeh[3] M. Askari[4]

1- Introduction

Investigations show that traffic collisions occur because of three main factors of human, road, and vehicle and proper planning and providing a solution for each can reduce the frequency of crashes, resulting in a decrease in road-related death tolls. Roads and road environments alone or in interaction with other factors (human and vehicle) account for 34% of total traffic collisions.

In order to provide effective solutions for road problems, the causes of incidents should be identified. In this regard, safety inspections and safety audits are very useful. The road safety audit process focuses on road safety and by identifying and correcting the elements which may lead to a collision play a significant role in increasing safety.

In this research, the aim is to provide a model for predicting crashes based on the results of the road safety audit. In this regard, eight factors affecting crashes include precipices vertical and horizontal signs, pavement situation , street gutters, roadsides and shoulder problems, ,access density, and road position which all are independent variables, and traumatic and fatal accidents are considered dependent variables.

According to the eight factors mentioned above, the road from Hamadan to Kermanshah, was divided into several parts and all the stages of road safety audits were passed through. Finally, considering the statistical relationships between the parameters, a model is devised that can be used to predict the likelihood of high-intensity crash occurrence, i.e., traumatic and fatal crashes, in a section of a road outside of a city

2- Methodology

In this study, Poisson regression and negative binomial regression models were used to model crashes. Description of these models is as follows.

Poisson regression model

In the statistics, Poisson regression is a kind of regression analysis and a subset of generalized linear models used to analyze the data obtained from counting. If the $x \in R^n$ is the vector of the dependent and independent, it will take the form of Eq. (1).

$$\log(E(Y \mid x)) = a'x + b \qquad (1)$$

If in eq. 1 $a \in R^n$ and $b \in R^n$, we can write Eq. (2) as follow:

$$\log(E(Y \mid x)) = \theta' x \qquad (2)$$

Where x is the n + 1 dimension of variables. By using the Poisson regression parameter and the input vector, the linear equation can be written and developed exponentially.

In the Poisson regression model, the kth value of the dependent variable y_k is modeled as a Poisson random variable with the mean μ_k and probability distribution function of Eq. (3):

$$p(y_k) = \frac{e^{-\mu_k} \mu_k^{y_k}}{y_k!} \qquad (3)$$

Negative binomial regression model

The most commonly used model for the fitting of bipartite counting data is a negative binomial, which is expressed in both first (NB-1) and second (NB-2) types, and for the data analysis, the constant and variable distribution problem is used, respectively. In the negative binomial regression model, the k-th observation of the dependent variable (y_k) has the probability distribution function of Equation (4):

$$p(y_k) = \frac{\Gamma(y_k + r)}{y_k!\Gamma(r)} \left[\frac{\mu_k}{\mu_k + r}\right]^y \left[\frac{r}{\mu_k + r}\right]^r \qquad (4)$$

The conditional mean for the vector of the observed variables (x_k) is obtained from equation (5):

$$E\left(y_k \middle| x_k\right) = \mu_k = e^{x'_{i\beta}} \qquad (5)$$

The relation between mean and variance in the distribution of negative binomials in the type of over-distribution of the variable (NB-2) is presented in Equation 6:

$$V\left(y_k\right) = \mu_k + \frac{1}{r}\mu_k^2 \qquad (6)$$

3- Modelling

Regarding the independent and dependent variables of the model shown in Table 1, several models for each model were constructed and their statistics and tests were compared. The best results of the negative binomial

[1*] Corresponding Author,Ph. D. Candidate, Tarbiat Modares University.
Email: m.asadamraji@modares.ac.ir
[2] Assistant Professor, Tarbiat Modares University.
[3] Professor, Tarbiat Modares University.
[4] Ph. D. Candidate, K. N. Toosi University of Technology.

regression model for injury and fatal accidents are shown in Tables 1 and 2, respectively.

Table 1. Output coefficients and output statistics of STATA software in the negative binomial model for injury accidents

Parameter	Coefficient	Z	P>|Z|
constant	2.38	8.96	0.000
I	-1.13	-2.61	0.009
N	-0.78	-1.92	0.055
Log likelihood: -180.45 LR chi2 (2):7.38 Pseudo R2: 0.02			

As Table 1 shows, in the best binomial model for injury crashes, two independent variables, precipices as well as the path smooth with a 99% confidence were found while the LR was 7.38.

Table 2. Output coefficients and statistics of STATA software in the negative binomial model for fatal accidents.

Parameter	Coefficient	Z	P>|Z|
constant	-0.24	-0.56	0.573
E	0.21	1.77	0.076
Log likelihood: -98.80 LR chi2 (2):3.26 Pseudo R2: 0.162			

According to the data in Table 2, in the negative binomial model of fatality crashes, the only independent significant variable with access reliability was 93%, and the LR was 3.26.

4- Conclusion

The results of this research are as follows:

- In models with different models and different combinations of independent variables, the best models were Poisson models of fatal accidents, as shown in Table 3.

Table 3. The variables with the most significant relationship between Poisson regression models

	Accident types (dependent variable)	Independent variables
1	Injury accidents	Access density, roadside problems, abyss, precipices , horizons and vertical arches
2	Fatality accidents	Access density, paving problems, abys, water problems and vertical arch

- Considering the results of this research, it could be stated that to reduce the number of casualties and injuries, priority should be given to enhancing safety in the area of precipices, fixing pavement problems, removing obstacles, removing barriers to vision, arch modification, and modifying access and their identification.