EVALUATING FACTORS AFFECTING THE EXPECTED TRAFFIC CRASH FREQUENCY BASED ON DATA FROM FARS

[1]MOHAMMED SAID OBEIDAT, [2]NADER AL THEEB, [3]RAHMA MOHAMMAD OBEIDAT

[1,2,3]Department of Industrial Engineering, Jordan University of Science and Technology, PO Box 3030, Irbid 22110, Jordan
E-mail: msobeidat1@just.edu.jo

Abstract –
The number of people using vehicles has increased dramatically. Traffic crashes are one of the leading causes of death around the world. In this study, crash data were obtained from the Fatality Analysis Reporting System (FARS) database for several years in The U.S., and were analyzed using the Statistical Analysis System (SAS©) software based on the negative binomial model. Acrash frequency modelwas constructed for nighttime crashes considering several factors. The significant factors in the nighttime crash frequency model included but were not limited to the area type (urban/ rural), the lighting condition, the speed related, the number of lanes, and the driver's age and gender.

Keywords - Roadway Safety; Count Data; The Negative Binomial Model; FARS

I. INTRODUCTION

Road traffic crashes are one of the main causes of death worldwide due to the dramatic increase in the number of people using vehicles. Traffic crashes increase traffic congestion on roads and boost medical costs (Lee & Li, 2014). Furthermore, traffic crashes are one of the leading causes of death and injuries in emerging countries (Hassouna & Tubaileh, 2021). Estimates by the World Health Organization (WHO) suggest that roadway crashes injure over 50 million and kill about 1.25 million people annually around the world (WHO, 2015). According to the China Annual Report for Road Traffic Accidents, crashes resultingin more than 10 fatalities are classified as extremely severe crashes (Xu, et al., 2020). Thus, an important strategy by decision makers is to reduce the crashes' impact and their economic and social impacts, precisely with severe crashes (Rezapour, et al., 2021).

Road traffic injuries are more dangerous and cause death than HIV/AIDS, tuberculosis, or diarrhoeal diseases (WHO, 2018). The number of road traffic deaths differs between countries, it depends on the income levels of countries (WHO, 2018). The number of deaths in low-income countries is 27.5 deaths per 100,000 populations, compared to that in high-income countries of 8.3 deaths per 100,000 populations (WHO, 2018). Vulnerable road users – motorcyclists, cyclists,and pedestrians, represent more than half of all road traffic deaths. About 26% of road traffic deaths are pedestrians and cyclists, and 28% are those using motorized two and three-wheelers. Car occupants represent 29% of all deaths, and the remaining 17% are unidentified road users (WHO, 2018).

Several factors contribute to crash frequency and severity. These are related to the crash, driver, vehicle, and road, along with the atmospheric factors(Delen, et al., 2017), (Ihueze & Onwurah, 2018). Previous studies showed that the most important factors that affect the crash were the driver's age, gender, collision type, crash traffic location, lighting, weather, road surface, and driving speed (Duddu, et al., 2019),(Wang & Zhang, 2017), (Obeidat, et al., 2022)in addition to the volume of the traffic, the density of the access point, vertical sinuosity, and lane width(da Costa, et al., 2018). Statistics showed that male drivers of ageless than 30 years or those above 70 years tend to beinvolved in severe crashes (Eustace & Wei, 2010). Bad weather, especially fog and rain, increases the crashes'severity and frequency (Wu, et al., 2018). In addition, dark road conditions (either light off or without lights) have a high opportunity of causing fatal crashes or severe injuries (Obeidat & Rys, 2017), (Wang & Zhang, 2017).

Previous studies analyzed crash data from a few states in the U.S. In this study, data from the 50 states in the U.S. will be considered. This paper aims to study the impact of several factors related to the driver, road, and vehicle on crash frequency. Data were collected from the Fatality Analysis Reporting System (FARS) for the years 2015-2018.The results of this study could improve the design considerations of vehicle manufacturers for the vehicle cabinet, to improve the safety of drivers and passengers.

II. OVERVIEW OF THE NEGATIVE BINOMIAL REGRESSION MODEL

The negative binomial regression is used when the dependent variable of interest measures the number of occurrences (counts), of some random event in a specified unit of space or time. Poisson regression is also an appropriate regression of count data, however, it does estimate that the mean is equal to the variance, which is not true in many cases. The negative binomial regression allows overdispersion, thus, if the researcher is not sure about overdispersion the use of the negative binomial regression is more appropriate (Cameron & Trivedi, 2013). In this study, negative

binomial regression is used to identify the factors that affect the crash frequency models.

The following equations or definitions for the negative binomial regression model are based on (Hilbe, 2011). In addition, the crash frequency model is used as an example to explain the equations.

For regression, it is assumed that:

$$y_i \sim negbin(\mu_i, k)$$
$$\mu_i = E(y_i) \quad (1)$$

Where $y_i$is the number of crashes at intersection i, and $E(y_i)$ is the expected crash frequency at intersection i.

The negative binomial density function is given by:

$$P(y_i) = \binom{k^{-1} + y_i^{-1}}{y_i} \left(\frac{k\mu_i}{1 + k\mu_i}\right)^{y_i} \left(\frac{1}{1 + k\mu_i}\right)^{\frac{1}{k}}, i = 1, 2, \dots, n \quad (2)$$

Where $k = 1/\alpha$ and α is the overdispersion parameter.

The link function of the negative binomial is:

$$\ln(\mu_i) = \beta^T X_i \quad (3)$$

Where μ_i is the expected crashes number at intersection i, β is the matrix of the estimated regression parameters, and X_i is the matrix of predictor variables.

The relation between the variance and the mean in the negative binomial distribution is:

$$\mathrm{Var}(y_i) = \mu_i + k\mu_i^2 \quad (4)$$

Where $\mathrm{Var}(y_i)$ is the variance of observed crashes y_i occurring at intersection i.

III. METHODOLOGY

Data were obtained from the FARS database. FARS provides statistics related to injuries and deaths from vehicle crashes in the 50 States of The U.S. The data include information relatedto the driver, the vehicle and the road for each crash.Data were analyzed statistically using SAS software based onlogistic regression. The negative binomial model was structured to analyze the collected count data.

The data were used to find the factors that affect the expected crash frequency. These data include several variables classified as independent variables and dependent variables. The independent variables were classified into driver characteristics, environmental characteristics, road characteristics, crash characteristics, and vehicle characteristics. One crash frequency model was constructed for nighttime crash data. In the crash frequency model, the count of crashes was the dependent variable that was used in this study.

The backward elimination procedure was used to estimate the significant variables based on a 5% significance level. The variable that has the highest p-value was removed first, followed by one of the next p-value, and so on until all the remaining variables were with a p-value less than or equal to 0.05.

IV. DATA

Data were obtained from the FARS database for all states from 2015 to 2018. Several factors were considered divided between those related to the driver, environment, road, crash, and vehicle. The driver-related factors included age, gender, injury severity, use of helmet, history of crashes, drink-driving, and hit-run crashes. The environmental factors included the weather and lighting conditions. The road factors included the area type (urban/rural), number of lanes, surface condition, national highway system, speed limit, intersection, road alignment, surface type, and roadway grid. The crash-related factors included the underride/override crash, rollover crash, speeding related, and extrication. The vehicle-related factors included vehicle type, hazardous material involvement, vehicle trailing, airbag, special use, and extent of vehicle damage.

V. RESULTS

Acrash frequency model was built for the obtained nighttime crashes in the years (2015-2018) for all states in The U.S.The Model was based on the negative binomial regression and showedthe frequency of crashes likelihood during nighttime. The full model of maximum likelihood parameters that estimate the frequency of crashes in all years that occurred during the night is shown in Table 1. The total number of crashes that were used in the statistical analysis was 61**,**148 **crashes**. Based on Table 1 and considering the significance level of 5%, the only insignificantvariable is the speed limit, as its p-value is 0.3197 which is greater than 0.05.

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	P-value
Intercept	1	-2.7171	0.1919	-3.0932	-2.3411	200.53	<0.0001
Area type (rural/urban)	1	-0.4059	0.0980	-0.5980	-0.2138	17.14	<0.0001
Lighting	1	0.3757	0.0886	0.2020	0.5494	17.97	<0.0001
Speed Related	1	0.9571	0.0811	0.7982	1.1160	139.35	<0.0001
Speed limit	1	-0.0811	0.0815	-0.2408	0.0786	0.99	0.3197
Number of lanes	1	1.5576	0.0944	1.3726	1.7427	272.26	<0.0001

Roadway alignment	1	1.1575	0.0833	0.9942	1.3207	193.13	<0.0001
Roadway grade	1	0.7816	0.0795	0.6258	0.9374	96.73	<0.0001
Surface type	1	3.3042	0.1161	3.0768	3.5317	810.63	<0.0001
Weather	1	0.8132	0.0772	0.6620	0.9645	111.03	<0.0001
Gender	1	1.2252	0.0787	1.0711	1.3794	242.67	<0.0001
Dispersion	1	0.8104	0.0461	0.7249	0.9061		

Table 1. Analysis of the maximum likelihood parameter that estimates the full nighttime crash frequency model.

Considering a 5% significance level and the backward elimination procedure, the reduced nighttime crash frequency model is shown in Table 2, where all included factors are significant, as all variables with a p-value less than 0.05, thus they were considered significant parameters. The significant variables of the model were speed-related, roadway surface type, roadway grade, area type, number of lanes, roadway alignment, light condition, gender, and weather conditions.

Parameter	**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**P-value**
Intercept	1	-3.0250	0.2183	-3.4528	-2.5972	192.05	<0.0001
Area type (rural/urban)	1	-0.3824	0.0968	-0.5722	-0.1927	15.60	<0.0001
Lighting	1	0.2850	0.0894	0.1097	0.4603	10.16	0.0014
Speed Related	1	1.0239	0.0851	0.8571	1.1908	144.68	<0.0001
Number of lanes	1	1.2034	0.0966	1.0140	1.3928	155.04	<0.0001
Roadway alignment	1	0.9700	0.0872	0.7991	1.1410	123.68	<0.0001
Roadway grade	1	0.7891	0.0837	0.6250	0.9531	88.86	<0.0001
Surface type	1	2.7061	0.1546	2.4030	3.0092	306.23	<0.0001
Weather	1	0.6088	0.0813	0.4494	0.7681	56.07	<0.0001
Gender	1	1.1016	0.0829	0.9391	1.2640	176.66	<0.0001
Dispersion	1	0.6717	0.0456	0.5880	0.7674		

Table 2.Analysis of the maximum likelihood parameter that estimates the reduced nighttime crash frequency model

VI. DISCUSSION

Based on the reduced nighttime crash frequency model and the estimated maximum likelihood parameters. The significant variables found were speed-related, roadway surface type, roadway grade, area type, number of lanes, roadway alignment, light condition, gender, and weather conditions, as their p-value, individually, is less than the 5% significance level.

For the area type (rural vs urban) the factor estimate was of negative value (-0.3824), indicating that the expected crashes during nighttime hours in urban areas are more than that in rural areas. This is explained by the traffic density in urban areas beingmore than that in rural areas. The lighting condition was also significant, with a positive estimate value (0.2850), indicating that night conditions without lighting increases the expected crash frequency due to bad vision. In addition, the speed-related factor was also significant with a positive estimate in the model, indicating that high driving speeds could increase the expected crash frequency during night hours. The number of lanes factor was also significant with a positive estimate in the model (1.2034), indicating that as the roadway lanes increases, the expected nighttime crash frequency increases. Moreover, roadway alignment was a significant factor with a positive estimate, indicating as the road curvature increases, the expected crash frequency increases. Furthermore, road grade also was a significant factor with a positive estimate, indicating that as the road levels increases, the expected crash frequency increases.

Regarding the driver gender, it had a positive estimate in the model, indicating that male drivers increase the expected crash frequency compared with female drivers. This is becausemale drivers are of a higher tendency to engage in risky driving behaviors, such as high speed and driving under influence of alcohol, etc.regarding weather conditions, which was a significant factor, with a positive estimate, it seems that the expected crash frequency increases in clear weather compared to bad weather conditions. This result could be due to many reasons. One of these reasons is the increased traffic jams in clear weather, and the divers'tendency to drive in less care clear weather compared to their driving tendency in bad weather, which is usually done with more care. Finally, based on the road surface condition, which

was significant and of positive estimate in the model the concrete, blacktop, bituminous, and asphalt surface types increase the expected crash frequency.

VII. CONCLUSION

In this study, data obtained from FARS were analyzed statistically based on the negative binomial regression model. A nighttime crash frequency model was constructed using data between 2015-2018. The result showed several significant variables that affect the expected crash frequency during nighttime. These were speed-related, roadway surface type, roadway grade, area type, number of lanes, roadway alignment, light condition, gender, and weather conditions.The results of this study may raise drivers' awareness of different causes of crashes that might affect their health.

REFERENCES

[1] Cameron, A. & Trivedi, P., 2013. Regression analysis of count data. 2 ed. United Kingdom: Cambridge University Press.

[2] da Costa, J. et al., 2018. Portuguese two-lane highways: Modelling crash frequencies for different temporal and spatial aggregation of crash data. Transport, 33(1), pp. 92-103.

[3] Delen, D., Tomak, L., Topuz, K. & Eryarsoy, E., 2017. Investigating injury severity risk factors in automobile crashes with predictive analytics and sensitivity analysis methods. Journal of Transport & Health, Volume 4, pp. 118-131.

[4] Duddu, V., Kukkapalli, V. & Pulugurtha, S. S., 2019. Crash risk factors associated with injury severity of teen drivers. IATSS Research, Volume 43, p. 37–43.

[5] Eustace, D. & Wei, H., 2010. The role of driver age and gender in motor vehicle fatal crashes. Journal of Transportation Safety & Security, Volume 2, pp. 28-44.

[6] Hassouna, F. & Tubaileh, M., 2021. Road traffic casualties in West Bank: Trends analysis and modeling. International Review of Civil Engineering, 12(2), pp. 101-107.

[7] Hilbe, J., 2011. Negative binomial regression. 2 ed. United Kingdom: Cambridge University Press.

[8] Ihueze, C. & Onwurah, U., 2018. Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria. Accident Analysis & Prevention, Volume 112, pp. 21-29.

[9] Lee, C. & Li, X., 2014. Analysis of injury severity of drivers involved in single- and two-vehicle crashes on highways in Ontario. Accident Analysis and Prevention, Volume 71, p. 286–295.

[10] Obeidat, M., Khrais, S., Bataineh, B. & Rababa, M., 2022. Impacts of roadway lighting on traffic crashes and safety in Jordan. International Journal of Crashworthiness, 27(2), pp. 533-542.

[11] Obeidat, M. & Rys, M., 2017. Comparing roadway lighting systems performance at rural and urban intersections. Pittsburgh, Industrial and Systems Engineering Research Conference.

[12] Rezapour, M., Wulff, S., Molan, A. & Ksaibati, K., 2021. Application of Bayesian ordinal logistic model for identification of factors to traffic barrier crashes: considering roadway classification. Transportation Letters, 13(4), pp. 308-314.

[13] Wang, Y. & Zhang, W., 2017. Analysis of roadway and environmental factors affecting traffic crash severities. Transportation Research Procedia, Volume 25, pp. 2119-2125.

[14] WHO, 2015. Global status report on road safety. Switzerland, World Health Organisation.

[15] WHO, 2018. Global status report on road safety 2018: Summary (No. WHO/NMH/NVI/18.20). s.l., World Health Organization.

[16] Wu, Y., Abdel-Aty, M. & Lee, J., 2018. Crash risk analysis during fog conditions using real-time traffic data. Accident Analysis and Prevention, Volume 114, pp. 4-11.

[17] Xu, C., Wang, C., Ding, Y. & Wang, W., 2020. Investigation of extremely severe traffic crashes using fault tree analysis. Transportation Letters, 12(3), pp. 149-156.

★★★