Logistic Regression Approach in Road Defects Impact on Accident Severity

Mohadeseh Khalili
Department of Statistics
University of Mashhad, International Branch
Mashhad, IRAN

Alireza Pakgohar
Department of Statistics
Payame Noor University (PNU)
Tehran. Iran
Email: a_pakgohar@pnu.ac.ir

***Abstract*— Safety and accident issues are considered as an important problem in the world. Road transportation issues would have a more conspicuous countenance in the country of Iran in which, over 94 percent of all transportations take place through roads. The roads' contribution to the traffic accidents at mean is 24 percent however this value is 36 percent in Iran. Road traffic crashes exert a huge burden on Iran's economy and health care services. Many parameters, safety signs and equipment, Vertical and horizontal arc combination along the road and all the effective factors in decreasing the accidents, are considered when determining the safety level of the roads. Road defect is being considered as one of the essential factors causing accidents to happen. This article investigates the severity factor of the accident according to the vehicle movability situation after the accident. The results of this research has shown that most important factors reducing the safety on the suburban roads in Iran is "insufficient road width" pertaining to frequency and "Level difference between road & shoulder" pertaining to accident severity.**

***Index Terms*— Road Defect; Iran; Accident Severity; Vehicle**

I. Introduction

The road accident issue is considered as an important problem in the world; To the extent that road accidents claim 1.2 million lives around the world each year and cause 50 million people to get injured [1, 2]. These statistical figures for Iran in 2007 were 27,567 casualties, 276,762 cases of injury and more than 442,000 crashes [3]. Iran, in terms of passenger car than the population at the end of 2007 was ranked forty-seventh in the world. Such that for every thousand people, 63 of which were passenger car in Iran [4].

The problem of accident is a very acute. Traffic accident leads to loss of life and property. Road accidents cannot be totally prevented but by suitable traffic engineering and management the accident rate can be reduced to a certain extent. For this reason systematic study of traffic accidents are required to be carried out. Proper investigation of the cause of accident will help to propose preventive measures in terms of design and control.

A comprehensive understanding of the relationship between road accident occurrence and severity of consequences permits the formulation of safety measures that are most cost-effective.

Factors that affect the level of damage experienced by individuals involved in road accidents include the following: accident dynamics, seating position, vehicle condition, vehicle size, driver condition, and driver action [5].

Road defect is being considered as one of the essential factors causing accidents to happen; which can include several items like traffic signs, asphalt and road surface, transverse inclination and road slope, arcs, road lighting, etc.

Regression models have a specific place amongst statistical techniques in recognize of effective factors in road accident severity [6, 8] for example the logistic regression models [9, 10, 11, 12, 13, 14]. Several studies using logistic regression pattern have been done in Iran, including national program for studies on causes and effective factors on road accidents, 2006 – 2007 [15] as well the studies about safety standards [16, 17].

II. Model and Methodology

Our model is mainly based on the following equation:

$$\mu(X,\beta) = P = \exp(\beta^T X)/1 + \exp(\beta^T X) \quad (1)$$

in which, β is the parameter vector. Especial:β_0 is the model constant and β_i are the parameter estimates for the independent variables.

The transformation of conditional mean $\mu(x)$ logistic function is known as the logit transformation. The Logit is the LN (to base e) of the odds, or likelihood ratio that the dependent variable is one, such that: Logistic regression is useful for predicting a binary dependent variable as a function of predictor variables. The goal of

doi:10.4304/jetwi.5.2.132-135

logistic regression is to identify the best fitting model that describes the relationship between a binary dependent variable and a set of independent or explanatory variables. The dependent variable is the population proportion or probability, P, that the resulting outcome is equal to one. Parameters obtained for the independent variables can be used to estimate odds ratios for each of the independent variables in the models [18, 19]. Let X, Y be a set of data with binary outputs. For each test x_i in X, the output can be y_i=0 or y_i=1. Our goal is to “Classify” the test xi based on a logistic regression model. Although *LR* is capable of handling a dataset with [0, 1] outputs, we limit our discussion to binary case. We use a Logistic Function to model the relationship between each x_i test and its expected output value. This function is written as:

$$logit(P) = LN(P_i/1 - P_i) = \beta_0 + \beta_i X_i \qquad (2)$$

In logistic regression, there is no true R2 value as there is in Ordinary Least Squares (OLS) regression [20]. In this study, there is Hosmer-Lemeshow statistics for goodness of fit recognizes [21]. Hosmer-Lemeshow (H-L) is the test statistic that obtained by applying a chi-square test on a $2\times g$ contingency table. The contingency table is constructed by cross classifying the dichotomous dependent variable with a grouping variable (with g groups) in which groups are formed by partitioning the predicted probabilities using the percentiles of the predicted event probability.

Classification Accuracy: It measures the proportion of correct predictions considering the positive and negative inputs. It is highly dependent of the data set distribution which can easily lead to wrong conclusions about the system performance [22]. It is calculated as follows Classification accuracy = Total hits / Number of entries in the set:

$$Accuracy = (TP + TN) / (P + N) \qquad (3)$$

Classification Sensitivity: It measures the proportion of the true positives, that is, the ability of the system on predicting the correct values in the cases presented. It is calculated using the following formula [22]. Sensitivity = Positive hits / Total positives:

$$Sensitivity = TP/(TP + FN) \qquad (4)$$

Classification Specificity: It measures the proportion of the true negatives, that is, the ability of the system on predicting the correct values for the cases that are the opposite of the desired one [22]. It is calculated as follows Specificity = Negative hits / Total:

$$Specificity = TN / (TN + FP) \qquad (5)$$

Data Usage: The data for analysis have been derived from reference (15). These data are based on the records available in Police Database (COM113 & COM 114) for the years of 2006 and 2007; which is including 279,432 records of accidents occurred on suburban roads. The entry method of logistic regression was followed using SPSS version 14 [23]. Our study in Iran’s road accident severity is the ruined vehicle immovable; In this regard, we would consider the state of vehicle road ability after the accident as an inactive factor for accident severity (able to movement & Non-able).

III. Results

Data availability has been the primary consideration in determining independent variables for this study. All variables mentioned in Table I were obtained from the police. This study attempts to consider all relevant road defects factors influencing vehicle road accidents. In addition Hosmer-Lemeshow (H-L) test shows the significance of the developed logistic regression models (*Sig*. > 0.05).

Based on the LR model results as shown in Table I indicate that at 99% confidence level, all independent variables were related to the ruined vehicle immovable after accident. The obstacles, Bumpers, and unknown defects (other) were negatively related to vehicle situation, and other independent variables were positively related. According to Table I, the possibility for the vehicle not being road able after the accident on an Iranian suburban road without any road defect is 25.6 percent. Maximum accident probability exists for the roads which does not have “Standard Level Difference between asphalt and shoulder”. The odd for the vehicle not being able to move after accident on these roads is 2.057 times more than on the roads without any safety defects. The second cause is non-standard “Transverse and longitudinal road slope”. On the roads having this defect, the odd for the vehicle not being able to move after accident is 1.975 times more than a vehicle that gets involved in an accident on a road without any safety defect. In other words, the probability for vehicle not being able to move after the accident on these roads is about 2 times more than on the normal suburban roads. The overall percentage of cases that are correctly predicted by the null model that is useful to quantify the diagnostic accuracy of the model is 76.7 based on null model. According to the Table II the sensitivity of final model in determining vehicle’s post-accident status (movability) is 50%. Besides the specificity measure is 83%. And more balanced accuracy is 66.5%. The final model correctly classifies 76.7% of cases in regard to vehicle post-accident movability status. The omnibus test of model coefficients are: x^2=3374.746, *d.f*=13, and *P-value* ≤ 0.001.

IV. Conclusion

The absolute frequency is not a suitable measure yardstick for determining priorities. As the results show in this study, insufficient road width has been contributing to more than 22% of suburban road accidents in Iran; But stood in the last rank as an accident severity cause factor. Our studies show that there is a linear correlation between driver death probability and vehicle inversion as two measurement factors for accident severity. Studies show that if some kind of road defect could cause any reduction of vehicle speed, it had reduced death probability for the driver furthermore.

Finally, suburban road accidents are considered as one most important concerns of health and safety authorities of the country. Iranian police has contrived many plans and paid enormous costs for driving behavior control to improve the safety belt usage from 60% to more than 95%. As a result the increasing accident rate has been reversed to a declining trend. However, the number of accidents has not declined in a tangible way, and number of casualties on the roads is still above 22 thousands. We believe there is a second reason to be considered that has been hidden due to paying excess attention to first reason; and that is crossover defects and low safety of the roads. Iranian roads contribute to more than 68% of the suburban accident cases. Generally, 36% of accident share in Iran is related to unsafe roads, while road defects in Europe contributed to 34% of the incidents and their share is 24% among three main factors according to Treit et al. report. Road safety defect cases on occurred incidents in Iran are two times more than Europe, and this fact shows that the second cause has been neglected to some extent. Many improvement strategies in Iran are based on further reduction in frequencies; while it is quite clear that the priorities should be pondered upon. It seems that there is a significant relationship between the severity of damage on the vehicle and the road defects. The null theory considered here is that there is no relationship between the damage severity on the vehicle and vehicle defects. The test theory is that such relationship exists between these two variables. Main road defect factors are examined and their relationship with accident severity is investigated in this section using regression equations.

TABLE I. REGRESSION MODEL COEFFICIENTS FITTED FOR VEHICLE STATE VARIABLE

Variable	Frequency	β	*Sig.*	Percent.	Odd	Rank
Non-standard transverse/longitudinal slope	1812	0.721	0.000	41.5%	2.057	1
Level difference between asphalt & shoulder	4117	0.681	0.000	40.5%	1.975	2
No guard rails	10267	0.674	0.000	40.4%	1.962	3
No soil shoulder & parking	7582	0.671	0.000	40.3%	1.955	4
Sharp angle bend	4503	0.518	0.000	36.7%	1.678	5
Asphalt surface defect	5215	0.359	0.000	33.1%	1.432	6
Horizontal signs defect	24277	0.174	0.000	29.1%	1.190	7
Road lighting defect	8114	0.135	0.000	28.3%	1.145	8
Vertical signs defect	27894	0.129	0.000	28.2%	1.138	9
Insufficient road width	61840	0.089	0.000	27.4%	1.093	10
Obstacles and bumpers	4255	-0.203	0.000	22.0%	0.816	11
Road depression	1392	-0.364	0.000	19.3%	0.695	12
Other	29719	-0.377	0.000	19.1%	0.686	13
Constant		-1.064	0.000	25.6%		

TABLE II. RANKING MATRIX OF VARIOUS MODELS FOR A GIVEN DATASET

Model ranking	**Observed**		
	model result	vehicle overturn	No vehicle overturn
Logistic Regression	vehicle overturn	68749	20075
	No vehicle overturn	68614	222564

ACKNOWLEDGMENT

Hereby the authors want to thank all the project team members who devotedly helped during this study. In addition, they would like to thank Arash Houshangi far for translation and documentation of the work. We are sincerely grateful to Traffic Police Practical Research Center for Supported by our studies.

REFERENCES

[1] Evans, L. Traffic safety. Bloomfield Mich. : Science Society, 2004.

[2] World Health Organization (WHO). Burden of disease project. Global burden of disease estimates for 2001. [Online] http://www3.who.int/whosis/menu.cfm?path=burden.

[3] Vahabzade, Ebrahim. Human Factor Effects on Driving Accidents on Karaj-Ghazvin Freeway on 2005, Solutions for Control and Decrease. Traffic Management Studies Quarterly. 2009. Vol. 8, Third year.

[4] http://niksalehi.com. [Online] Niksalehi, April 5, 2007. [Cited: August 11, 2011.] http://forum.niksalehi.com/showthread.php?t=16500.

[5] S. A. Nassar, F. F. Saccomanno, J. H. Shortreed, Selection of model structure for road accident severity. Canadian Journal of Civil Engineering, 1995, 22(5): 981-991, 10.1139/l95-114.

[6] The Effect of Visual Field Defects on Driving Performance; A Driving Simulator Study. Tanja R. M. Coeckelbergh et al. s.l. : Arch Ophthalmol, November 2002, Vol. 120 No. 11.

[7] Ground-borne vibration generated by vehicles crossing road humps and speed control cushions. G. R. Watts and V. V. Krylov. 3, s.l. : Applied Acoustics, March 2000, Vols. 59, pages 221-236.

[8] Washigton, S p, Karlaftis, M G and Mannering, F L. Statistical and Econometric Methods for Transportation Data Analysis. s.l. : Chapman & Hall, 2003.

[9] Herman F. et al. Evaluation of Lane Reduction "Road Diet" Measures and Their Effects on Crashes and Injuries.

Transportation Research Record. FHWA-HRT-04-082 , HRDS-06/3-04(1.7M)E, 1784.

[10] An application of a conditional logistic regression to study the effects of safety belts, principal impact points, and car weights on drivers' fatalities. Kung-Jong Lui et al. 4, s.l. : Journal of Safety Research, Winter 1988, Vols. 19, p.p. 197-203.

[11] An integrated approach for studying the safety of road networks: logistic regression models between traffic accident occurrence and behavioural, environmental and infrastructure parameters. F. Crocco et al. s.l.: WIT Transactions on Ecology and the Environment © WIT Press doi, 2010, Vol. 142. ISSN 1743-3541.

[12] The involvement of drugs in drivers of motor vehicles killed in Australian road traffic Crashes. Olaf H. et al. 2, s.l.:Accident Analysis & Prevention, March 2004, Vols. 36, pages 239-248.

[13] Combining non-parametric models with logistic regression: an application to motor vehicle injury Data. Petra M. et al. 3, s.l. : Computational Statistics & Data Analysis, 28 September 2000, Vols. 34, pages 371-386.

[14] The Risk of Dying in Alcohol-Related Automobile Crashes among Habitual Drunk Drivers. Robert D. Brewer et al. s.l. : The New England Journal of Medicine, 1994, Vols. 331, p.p. 513-517.

[15] Pakgohar, Alireza. An examination of causes and effective factors on reducing traffic accidents, based on CART and LR regression. s.l. : Applied Research Office, Guidance and Driving Police, 2007.

[16] The Effect of Visual Field Defects on Driving Performance; A Driving Simulator Study. Tanja R. M. Coeckelbergh et al. s.l. : Arch Ophthalmol, November 2002, Vol. 120 No. 11.

[17] Ground-borne vibration generated by vehicles crossing road humps and speed control cushions. G. R. Watts and V. V. Krylov. 3, s.l. : Applied Acoustics, March 2000, Vols. 59, pages 221-236.

[18] Washigton, S p, Karlaftis, M G and Mannering, F L. Statistical and Econometric Methods for Transportation Data Analysis. s.l. : Chapman & Hall, 2003.

[19] Pakgohar, Alireza. Statistics in Traffic Management. Tehran : Police University (Oloome Entezami), 2009. 978-964-7648-69-1.

[20] Emergency Trip Destination of Evacuation as Chelter Analysis for Tsunami Disaster: A Case Study in Phuket, Thiland. Charnkol, T, Hanaoka, S and Tanaboriboon, Y. s.l. : Jurnal of the Estern Asia Society for Transportation Studies, 2007, Vol. 7, pp. 853-867.

[21] Hosmer, David W and lemeshow, Stanley. Applied Logistic Regression.s.l. John Wille, 2000. 0-471-35632-8.

[22] Pujari, Pushpalata. "Classification and Comparative Study of Data Mining Classifiers with Feature Selection on Bionomial Data Set", Journal of Global Research in Computer Science 3.5 (2012): 39-45.

[23] Brace, Nikola, et al. SPSS for Psycologists: a guide to data analysis using. Tehran : Dowran, 2009. 978-964-8891-77-5

.