

# A Study of Structure of Road Accidents Status Prediction Model for Highways

Hassan Zoghi, Morteza Tolouei, Kianoush Siamardi, and Mohammad Ramezani Afjadi

**Abstract**—With regard to the importance of road accidents and the resulting costs, which include human, economic and social losses, the study and modeling of road accidents will be inevitable in identifying the causal factors of accidents. This article analyzes the construction of a model for the severity of fatal road accidents taking into account of the characteristics of the driver, motor vehicle, road and the environmental conditions of the roads of the country. In order to construct the accidents severity model, the Logit function of fatality likelihood is used. The function's output varies between zero and one. As the output becomes closer to one the likelihood of fatality will be more, and as it approaches the zero the likelihood of fatality will come down. In such models, such factors as elderly drivers, the existence of motorcycles and bicycles, inattention to the front, exceeding the authorized speed, unauthorized deviation to left due to overtaking, passing unauthorized place, illiteracy of drivers, etc. will increase the severity of accidents, and such factors as clear weather, standard road width, etc. will decrease fatal accidents. The factors are discussed in detail in this article.

**Index Terms**—Road accident, highways, logit model, accident indexes.

## I. INTRODUCTION

The expansion of the road transportation network of the state and increased volume of commutes, and an overview of the road accidents during the recent years, suggest need for more research on accidents. According to the statistics, more than 14,000 people per year lose their lives in road accidents and close to 30,000 people become injured and disabled, so that the drivers of motor vehicles account for 36 percent of the people killed in highways accidents; the occupants account for 45 percent of those killed; and pedestrians constitute 19 percent of the fatalities [1]. Of the whole roads of the country, 0.9 percent are freeways, 31.1 percent major roads, and 49.7 percent secondary roads, and 18.3 percent include are roads within jurisdiction of Ministry of Road and Transportation, so that more than 70 percent of the whole accidents take place on freeways and major roads. An

analysis of the highways accidents in Iran suggests that the design factors contributes to the accidents by 2 percent, motor vehicle factor by 13 percent, road factor by 22 percent, and human factor by 63 percent [2]. Therefore, with regard to the importance of the road accidents the need for research into the subject seems of great importance.

## II. PREVIOUS STUDIES

In 1998, J'wyn studied the number of accidents in the four-lane expressway in New Jersey. His studies showed that a highest percentage of accidents happened when the rate of commute was at its lowest level. Further studies by Change led to achievement of U-shaped models. The studies suggested how drivers reacted to compound or complicated traffic signs and systems and change their driving behaviors in different roads [3].

Madanat et al. developed binary logit models for likelihood prediction of two types of freeway incidents: accidents and overheating vehicles. They considered both loop and environment data in their model development. They found peak period, temperature, rain, speed variance, and merge section to be significant predictors of overheating incidents. For the crash prediction model, only three variables were found to be statistically significant, which are rain, merge section, and visibility [4].

Another promising research direction involves combining mechanism and statistical models. That is, where plausible causal variables can be identified, but where the actual form of the structural equation remains elusive, a statistical model might be used to approximate it. A good example is the work by Kloeden et al, where accident reconstruction methods were used to estimate speeds of accident-involved vehicles, and these speed estimates were in turn used to estimate a logit model relating speed to accident risk [5]. The logit model was then used to estimate how the probability of collision might change as speed was reduced. A Bayesian version of this approach was later described in Davis et al [6]. Hourdos used logit modeling to test for conditional independence in the construction of a detailed graphical causal model for rear-ending accidents, on a section of urban freeway [7].

## III. METHOD OF ANALYSIS

### A. Random Desirability Theory

The desirability function is used as a random characteristic to show the uncertainty and the basic presumption is the decider's complete ability to make a distinction. Manski identified and introduced four different sources for

Manuscript received February 18, 2012; revised March 30, 2012. This paper was supported by the Islamic Azad University, Karaj Branch.

Hassan Zoghi is with the Civil Engineering Department, Islamic Azad University, Karaj branch, member of IACSIT, Tehran, Iran (e-mail: H\_zoghi@kiau.ac.ir).

Morteza Tolouei is with the Tehran Industrial Cities (Iran Ministry of Industry, Mine, Trade) as Technical Expert (e-mail: morteza.tolouei@gmail.com).

Kianoush Siamardi is with the Atisaz Construction Co, Tehran, Iran (e-mail: k.siamardi@gmail.com).

Mohammad Ramezani Afjadi is with Obermeyer Middle East GmbH, He is M. Sc Road and Transportation Assistant, Abu Dhabi (e-mail: mohammad.ramezani@obermeyer.ae).

uncertainty in 1977 [8], in spite of lack of information and uncertainty as a result, that are:

- Failure to take account of certain attributes or options,
- Failure to take account of individuals differences
- Existence of measurement-related errors

Failure to question directly and obtain information from a certain individual or group on behalf of others.

The desirability of choosing by a typical person  $n$  of the option  $i$  out of a set of options  $C_n$  is defined as follows [9]:

$$U_{in} = V_{in} + \varepsilon_{in} \quad (1)$$

$V_{in}$  is the deterministic part of desirability,

$\varepsilon_{in}$  is the random statement of desirability (suggesting the uncertainty),

The deterministic statement of the model may be considered a function of the specifications  $X$ , written as:

$$\sum \theta_{Ki} X_{ikn} = V_{in} \quad (2)$$

It is normally presumed that the parameter  $\theta$  is constant for all people although it can take various values for different options.

The option which has the highest desirability is chosen. Therefore, the probability that the option  $i$  of the set of options  $C_n$  is chosen by the decider  $n$  is:

$$P(i|C_n) = P[U_{in} \geq U_{in} \forall j \in C_n] = P[U_{in} = \max U_{in}] \quad (3)$$

#### B. Logit Model

Logit model is a mathematical method which may be used to describe the relation between several variables  $X$  and one dependent variable with two (and more) dimensions,  $D$ . The function used in this method is an S-shaped function named Logit function whose domain is infinite and its range varies between zero and one. The mathematical formula of the function is as follows:

$$\begin{cases} f(y=1) = \frac{1}{1+e^z} \\ Z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k = \alpha + \sum \beta_i X_i \end{cases} \quad (4)$$

where the variable  $D$  is a variable with the two values zero and one or the two values B and C. Writing the likely model of interest as a conditional probability will give:

$$\begin{cases} P(D=1|X_1, X_2, \dots, X_k) = \frac{1}{1 + \exp(\alpha + \sum \beta_i X_i)} \\ P(D=0|X_1, X_2, \dots, X_k) = 1 - P(D=1|X_1, X_2, \dots, X_k) \end{cases} \quad (5)$$

The method by which the unknown parameters of the model, that is  $\alpha, \beta_1, \beta_2, \dots, \beta_k$ , is referred to as “maximal likelihood method”.

In Logit model, the variable  $D$  may be defined as a multi-dimension or multi-value variable. Consider for example a certain individual has the option to choose his mode of travel out of the three modes personal motor vehicle, shared motor vehicle and bus. If the three modes are shown by three characters 0, 1 and 2, at its easiest, by considering one of the three modes (say, zero) as the basis one can write [9]:

$$\begin{cases} P(D=0) + P(D=1) + P(D=2) = 1 \\ P(D=1) + P(D=0) \neq 1 \\ P(D=2) + P(D=0) \neq 1 \end{cases} \quad (6)$$

$$\begin{cases} P(D=0|X) = \frac{1}{1 + \exp[h_1(X)] + \exp[h_2(X)]} \\ P(D=1|X) = \frac{\exp(h_1(X))}{1 + \exp[h_1(X)] + \exp[h_2(X)]} \\ P(D=2|X) = \frac{\exp(h_2(X))}{1 + \exp[h_1(X)] + \exp[h_2(X)]} \end{cases} \quad (7)$$

The generalization of the Logit model to more than two alternatives has been known as “polynomial Logit model”. Polynomial Logit model has been taken from the presumption that the error statements of desirability functions are independent and follow the Gumbel distribution in the same way. The error statement of  $\varepsilon_{in}$  for every  $i$  and  $n$  is distributed as follows [9]:

$$\begin{cases} F(\varepsilon) \exp[-e^{\mu(\varepsilon, \eta)}] \mu > 0 \\ f(\varepsilon) \mu e^{\mu(\varepsilon, \eta)} \exp[-e^{\mu(\varepsilon, \eta)}] \end{cases} \quad (8)$$

where  $\eta$  is the place parameter and  $\mu$  is a strictly positive parameter referred to as the scale parameter. The average of the distribution is  $\gamma/\mu + \eta$  where

$$\gamma = \lim_{K \rightarrow \infty} \sum_{i=1}^K \frac{1}{i - \ln(K)} \quad (9)$$

is Euler’s constant. The variance of the function is  $\pi^2/\sigma\mu^2$  [9].

#### C. Analysis of Logit Model Sensitivity

The analysis method of Logit model is of the type maximal likelihood and the method of analysis of the model defines the type of the assessment rules. The method is based on maximizing the likelihood that the events happen at the same time and the related function is referred to as “log-likelihood function”. In this method, the likelihood of every observation is reckoned on the basis of the model to be used according to the model’s parameters. The likelihood that the observations take place at the same time is given by multiplication of the likelihoods of all the individual events. The function is referred to as the “function of likelihood” and is as follows [3]:

$$L(\beta) = \prod_{i=1}^N P_i \quad (10)$$

where  $L(\beta)$  is the function of likelihood of the observations 1 through  $n$  for the  $\beta$  coefficients.  $P_i$  is the likelihood of observation.

The method is intended to find  $\beta$ ’s that render the function maximal. In other words,  $\beta$ ’s should be sought which maximize the likelihood that the observations happen at the same time.

The goodness of fit is a criterion which defines the model’s

degree of goodness of fit for the data and its value varies between zero and one. The goodness of fit for Logit models is defined as follows [10]:

$$\rho^2 = 1 - \frac{LL\beta}{LL(o)} \rho_c^2 = 1 - \frac{LL\beta}{LL(\alpha)} \quad (11)$$

where  $\rho^2$  is goodness of fit  $\rho_c^2$  is goodness of fit  $LL(o)$  is the value of the log-likelihood function for zero coefficients.

$LL(\alpha)$  is the value of the log-likelihood function for constant statements.  $LL(\beta)$  is the value of the log-likelihood function for estimated coefficients.

#### IV. STATISTICAL SAMPLES FOR MODEL CONSTRUCTION

Normally, in order to study the extent different factors contribute to the severity of road accidents, the road accidents databank of State Organization for Transportation and Terminals is used, which includes the accident information that is completed and gathered on the spot and is available as a DBF (database file). The statistics gathered relate to the main roads of the highways, which may be easily used in the model's structure [2].

#### V. DESCRIPTION OF VARIABLES

For the purpose of study and modeling of the severity of accidents, changes are made to the databank's structure and the data available are converted to nominal variables that denote the mode of interest does or does not exist. The dependent variable in these analyses is the different levels of severity of accidents (fatal accidents, accidents leading to injury, and accidents causing only damage), and the independent variables include the human factors, motor vehicle, road and environmental conditions. As there are small numbers of accidents causing only damage and the accidents leading to injury and those causing only damage are consolidated, the levels are accidents are normally divided into two classes of accidents: accidents causing injury and fatal accidents.

#### VI. CONSTRUCTION OF MODEL

As the dependent variable is the severity of accidents and finally will be defined as fatal and leading to injury, in the analyses using Statistical Package for Social Science (SPSS) software the fatal accidents are denoted by figure zero and those leading to injury by figure one. In the model, the sensitivity analysis is conducted by  $LL(\beta)$  or log-likelihood function, Significant level (SIG) of the model's variables coefficients and the correct percentage, which suggests the percentage of observations that the model predicts correctly.

Normally, the model's assessment is obtained by the variables significance with 95% of certainty ( $\text{sig} < 0.05$ ) and higher percent correct and less absolute value ( $LL(\beta)$ ). The coefficients obtained relate to the function of death likelihood, and the variables with negative coefficients suggest a rise in the severity of accidents and variables with positive coefficients mean a fall in the severity of accidents. To construct the models of the accidents status prediction for

highways, variables becoming significant by step-by-step method given by the relations 4 through 13 in this article are used.

TABLE I: VARIABLES BECOMING SIGNIFICANT BY STEP-BY-STEP METHOD

Variable No.	Variable description	Variable name
1	Middle-aged drivers 1; otherwise 0	DA
2	Elderly drivers 1; otherwise 0	DA
3	Motorcycles and bicycles 1; otherwise 0	VEH
4	Inattention of the front 1; otherwise 0	AR
5	Exceeding the authorized speed 1; otherwise 0	AR
6	Deviation to left due to overtaking 1; otherwise 0	AR
7	Wrongful turns 1; otherwise 0	AR
8	Violation of no entry sign 1; otherwise 0	AR
9	Illiterate 1; otherwise 0	DG
10	Other cases 1; otherwise 0 (human cases)	HF
11	Intersection 1; otherwise 0	LS
12	Constant	CON

The result of fitting of the final model is obtained on the basis of Table II. The resulting  $\rho^2$  value for the final model suggests a proper or improper status of fitting the model to the data.

TABLE II: RESULT OF FINAL FITTING OF MODEL GIVEN BY THE FOLLOWING VALUES MEASUREMENT

Number of observations	N
Value of log-likelihood function for a zero coefficient	$LL(o)$
Value of log-likelihood function at convergence point	$LL(c)$
Value of log-likelihood function for constants	$LL(\beta)$
oodness of fit coefficient	$\rho^2$

#### VII. CONCLUSION

Considering the studies conducted, the results of the studies in the roads of the country suggest that such factors as the elderly drivers, motorcycles and bicycles, inattention of the front, exceeding the authorized speed, deviation to left due to overtaking, wrongful turns, entry to no entry streets, illiteracy of drivers and the existence of intersections will increase the severity of road accidents, and such factors as middle-aged drivers, clear weather, standard road widths, etc will decrease the severity of fatal accidents. The parameters and variables capable of measurement are given in Table I and the model's final fitting method is given in Table II.

#### ACKNOWLEDGMENT

It is appreciated from Islamic Azad University, Karaj Branch for preparing this paper which has been extracted from research study of mentioned university.

#### REFERENCES

- [1] *An Analysis of Highway Accidents*, Statistics and Planning Office, State Organization for Transportation and Terminals, safety of ways

commission, *Iran ministry of road and transportation instruction*, 2007, pp. 12-20.

- [2] *Statistics of Highway Accidents*, Transportation Industry Magazine, Comprehensive Manual of Iran and the World Transportation Industry, Tehran, 1997.
- [3] T. Garling, T. Laitila, and K. Westin, "Theoretical Foundations of Travel Choice Modeling," Pergamon Press, 1998, pp. 15-80.
- [4] S. Madanat, P. Liu, H. Pan, and H. Teng, "Freeway Incident Likelihood Prediction and Response Decision-Making," In *Joint Highway Research Project Report FHWA/IN/JHRP-96/6*, Project No. HPR-2090, Purdue University, West Lafayette, Indiana, 1997.
- [5] C. Kloeden, J. McLean, and G. Glonek, "Reanalysis of Traveling Speed and the Risk of Crash Involvement in Adelaide South Australia," *Road Accident Research Unit*, University of Adelaide, 2002.
- [6] G. Davis, J.P. Pei, and S. Davuluri. (2006). Speed as a risk factor in serious run-off-road crashes: Bayesian case-control analysis with case speed uncertainty. *Journal of Transportation and Statistics*, vol. 9, pp. 17-28.
- [7] J. Hourdos, "Crash-Prone Traffic Flow Dynamics: Identification and Real-Time Detection," Doctoral Thesis, *Dept. of Civil Engineering*, University of Minnesota, 2005.
- [8] C. Manski, *the Structure of Random Utility Models, Theory and Decision*, 1977, ch. 8, pp. 229-254.
- [9] M. Ben-Akiva and S. R. Lerman, "Discrete Choice Analysis: Theory and Application to Travel Demand," MIT Press, Cambridge, MA, 1985.
- [10] J. N. Ivan, J. Paul, O. buggen, X. Qin, and J. Penda, "Pedestrian Crash Rate: Alternative Measures of Exposure," *University of Connecticut and University of New Hampshire*, 2000.



**Hassan Zoghi** was born in Tehran in 1966. He could achieve Traffic and Transportation Engineering PHD from Iran University of Science and Technology in 1998. His specialization is Traffic and Transportation, IT, ITS and Construction.

He is Assistant Professor of Civil Engineering Department of Technical Faculty, Islamic Azad University (Karaj Branch, South Tehran Branch, Science and Researches Branch). He is teaching in 9 Academies and schools (1993- hitherto). He is advisor and guide professor in more of 40 thesis. Additional, he is author of 45 papers in the conferences, seminars and journals and 2 books titled road design and road pavement. Dr. Zoghi is manager of IT in society of Iran traffic elites (at present) and other of executive profession (2001-2010). He is member of IACSIT. He has two papers which were indexed in conference proceedings citation index (Thomson REUTERS).



**Morteza Tolouei** was born in Tehran in 1982. He has graduated by degree of B.S.C of civil engineering from Islamic Azad University, Karaj Branch, Tehran, Iran in 2008. His major field of study is traffic engineering and pavement retrofitting. His specialization field is road and transportation, construction.

He is Technical Expert of Tehran Industrial Cities (Iran Ministry of Industry, Mine, Trade). He was

Project manager of RCC pavement of Nasirabad industrial city(Tehran). Additional, he was assistant project manager of pavement - shahriar industrial city and assistant project manager of network collect of surface waters - Nasirabad industrial city. He has published more than 30 papers in journals, international and national conferences.

Mr. Tolouei is member of young researchers club, Iranian concrete institute and organization for engineering order of building –Tehran. He has two papers which were indexed in conference proceedings citation index (Thomson REUTERS).



**Kianoush Siamardi** was born in Bandar Abbas Southern Iran in 1985. He has graduated by degree of B.S.C of civil engineering from Islamic Azad University, Karaj Branch, Tehran, Iran as excellent student in 2008. His major field of study is traffic engineering and pavement retrofitting. He worked on many research topics such as traffic safety, road design, ITS, concrete pavement, seismic retrofitting of RC structures. Already he is working as Manager of Concrete Research Lab of Atisaz Construction Co and Supervisor of Resin-Epoxy Injection Project. So far, he has published more than 55 papers in journals, international and national conferences such as 11th IEEE ITSC 2008, 2009 ICCAE and etc.

Mr. Siamardi is member of IACSIT, young researchers club, Iranian concrete institute and organization for engineering order of building –Tehran. He gained excellent researcher student of Islamic Azad University, Karaj branch in 2008. In addition, he could earn first grade of ACI scholar competition in RC bowling ball in 2007. He has two papers which were indexed in conference proceedings citation index (Thomson REUTERS).



**Mohammad Ramezani Afjadi** was born in Esfahan in 1980. He was graduated by M.Sc. in Civil engineering (Road and Transportation in 2005, from Islamic Azad University, South Tehran branch, Iran. His specialization is Project Management, Construction Coordination, Road and Airport Design, QA/QC. He has worked on many executive civil project such as Semiconductor Manufacturing (2011), Abu Dhabi, Special Security Hospital (2010), Abu Dhabi, HH Sheikh Khalifa Bin Zayed Al Nahayan Hospital(2010), Homs, Syria, Paris-Sorbonne University(2008-2011), Abu Dhabi, Bypass Pipe Network, from Greens Wastewater Treatment Plant to IRR2 Lake, Emirates living development, design and build(2007-2008) Dubai, as a Member of Design Team in Infrastructure Section. He was construction project coordinator- Planning Engineer of many projects such as Persian Gulf International Airport. Bushehr, Assaluyeh port, Iran (2004-2007), 450 units of residential buildings, 9 towers, 14th Floor, BAME TEHRAN (Saadat abad), Iran (2003-2004). Already he is M.Sc Road and Transportation Assistant of Obermeyer Middle East GmbH .

He is member of organization for engineering order of building –Tehran.