# Crash Severity Modeling in Urban Highways Using Backward Regression Method

F. Rezaie Moghaddam, T. Rezaie Moghaddam, M. Pasbani Khiavi, M. Ali Ghorbani

*Abstract*—Identifying and classifying intersections according to severity is very important for implementation of safety related counter measures and effective models are needed to compare and assess the severity. Highway safety organizations have considered intersection safety among their priorities. In spite of significant advances in highways safety, the large numbers of crashes with high severities still occur in the highways. Investigation of influential factors on crashes enables engineers to carry out calculations in order to reduce crash severity. Previous studies lacked a model capable of simultaneous illustration of the influence of human factors, road, vehicle, weather conditions and traffic features including traffic volume and flow speed on the crash severity. Thus, this paper is aimed at developing the models to illustrate the simultaneous influence of these variables on the crash severity in urban highways.

The models represented in this study have been developed using binary Logit Models. SPSS software has been used to calibrate the models. It must be mentioned that backward regression method in SPSS was used to identify the significant variables in the model.

Consider to obtained results it can be concluded that the main factor in increasing of crash severity in urban highways are driver age, movement with reverse gear, technical defect of the vehicle, vehicle collision with motorcycle and bicycle, bridge, frontal impact collisions, frontal-lateral collisions and multi-vehicle crashes in urban highways which always increase the crash severity in urban highways.

*Keywords*—Backward regression, crash severity, speed, urban highways.

## I. INTRODUCTION

ALTHOUGH considerable efforts have been made to investigate the severity of traffic crashes, the relationship between risk factors and severity of crashes hasn't been properly identified. One of the reasons may be that the factors associated with levels of crash severity are complicated due to a number of factors including personal characteristics such as age, vehicle such as type of vehicle, environmental such as weather condition and road such as geometrical design [1].

So, in this study, a mathematical model capable of

F. Rezaie Moghaddam, Department of Engineering, University of Mohaghegh Ardabili, Ardabil, Iran (Corresponding author, Phone: +98-451-5512804; fax: +98-451-5512904; e-mail: F_Rezaie@uma.ac.ir).

T. Rezaie Moghaddam, Sama Unit of Azad University of Ardabil, Ardabil, Iran ( e-mail: Tagi_Rezaie@yahoo.com)

M. Pasbani Khiavi, Department of Engineering, University of Mohaghegh Ardabili, Ardabil, Iran (e-mail: pasbani@uma.ac.ir).

M. Ali Ghorbani, Department of Engineering, University of Mohaghegh Ardabili, Ardabil, Iran (Corresponding author, Phone: +98-451-5512804; fax: +98-451-5512904; e-mail: gorbani@uma.ac.ir).

suggesting the simultaneous effects of human factor, road, vehicle, weather condition and also traffic characteristics such as volume and flow speed on the severity of crashes will be presented.

By investigating the site of crash and the information and conditions related to the crash, a relationship between factors involved in crash occurrence and severity, for example fatalities, injuries or financial loss, will be presented in models defining the severity of crashes. So, such an approach will help in identifying the major factors related to crash severity and then taking countermeasures [2]. In these models, the variation of human, environment, geometrical design, vehicle and traffic factors are used.

The crash severity models are of great importance because in these models the crash-related factors, particularly fatalities, are identified and effective preventive measures are prioritized.

## II. BACKGROUND

Chen and Jovanis obtained a relationship between crash severity and its associated factors using Loglinear model. They used 408 observations of bus crashes in a freeway in Taiwan during of 1985 to 1993. They considered the crashes fatalities with crash injuries because the crashes fatalities were with low number. Chen and Jovanis emphasized on the importance of proper classification of a number of data such as the time of crash. Frontal impact collisions, driving in late hours of night or early morning and driver fault were introduced the main factors which have considerable affects on crash severity [3]. Kockelman investigated on models of crashes including two-vehicle, single-vehicle and all other types of crashes, separately. Because of the different nature of crashes and their cause, it is suggested to separate them in order to get the best results in the model. Finally, he identified the head-on impact collision, high speed, rollover, alcohol use, older age, overtaking maneuver, night crashes, etc as high severity crashes and rear or lateral impact collisions and day crashes as low severity ones [4]. Beside explaining and presenting the importance of crash severity models, Saccomanno suggested that the validity of current models is generally based on the accuracy of provided data from Reference databank. He used the Binary Logit models in his investigations and considered factors including driver fault, bad conditions of the driver (being tired or sick), poor vision, wet road, night crashes, vehicle breakdown, alcohol use, etc as increasing factors of crash severity and factors such as belt seat usage as decreasing one. In representing models,

Saccomanno showed that if in a category (for example fatality crashes), the number of observations were less compared with total crashes, its combination with injury crashes and considering both as one category will lead to more meaningful variables and better model results. Voget and Bareed obtained a relationship between rural two-lane road crash severity and its contributing factors [5]. Kim and Nitz investigated the relationship between crash type and severity using Loglinear model. They used crash data forms that have been completed by police at the site of the crash in Hawaii State. Finally, they considered head-on crashes and rollover as most severe ones [6]. Besides emphasizing on the effects of weather conditions and crash type on crash severity, Khattak defined factors such as careless driving, exceeding speed, alcohol use, younger or older drivers (<25 or >25), wet road, road turns, road grade, etc as increasing factors of crash severity and snow, day hour crashes, smooth road, high traffic volume, etc as decreasing ones using ordered Probit model [7]. In prior studies using artificial neural networks as the modeling approach for crash severity and it's relating factors such as human factors, roadway condition, weather condition and etc, has been scarce. Abdelwahab and Abdel-Aty classified crash severity into three injury severity levels, that is related to two-vehicle accident that occurred at signalized intersections. They used MLP for classifying data and get correctly classifying 65.6% and 60.4% of cases for the training and testing data, respectively [8].

Abdelwahab and Abdel-Aty also applied multilayer perception and fuzzy adaptive resonance theory to analyze driver injury severity in traffic accidents. The results indicated that gender, vehicle speed, seatbelt use, vehicle type, point of impact and area type of accident location can affect in probability of injury severity [9]. Dursun Delen et al. developed eight binary neural models to classify accidents by level of injury severity from no-injury to fatality and conducted sensitivity analysis to identify the prioritized importance of crash-related factors [10].

Generally, literature review represents different results on the effects of influential factors associated with crash severity. Actually, in anyone of these investigations, the increasing or decreasing factors of crash severity identified considering available data type.

## III. DATA COLLECTION

Developing a model requires some data and information, so in the current study, highways of Tehran (capital of Iran) has been considered as a case study and the required information were collected in the following steps:
1) Gathering information related to traffic volume and speed in different sections of urban highways.
2) Collecting information related to geometrical characteristics such as shoulder width, lane number and width in different sections of urban highways.
3) Selecting data related to urban highway crashes in recent years.

In order to investigate the influence rate of various factors such as environment, traffic, human, geometrical feature, vehicle, etc on crash severity in highways, data associated with traffic characteristics such as volume and flow speed and highways geometrical characteristics collected from Tehran Traffic and Transportation organization and Tehran comprehensive Traffic and transportation studies company, respectively. Information related to crash, obtained from Tehran police department databank, prepared from gathering completed forms by police at the site of the crash (Form "k113"). Regarding the qualitative nature of the information inserted in the form, data should be defined quantitatively. It is identified that the best method was to define variables as 0 and 1 for data analyzing and usage. Then, obtained information was delivered into Access software. It must be mentioned that for data and information to be used in modeling they should be placed in a databank in which the information collected from separate organizations and companies, could be investigated and those information that over lap with one another and also those with probable error could be identified and excluded from the databank in Access software.

Finally, data collected from crashes (regarding a number of problems associated with preparing this type of information through competent sources), traffic and geometrical characteristics for a 4 year period during 2004-2008 (regarding limitations in information resources) includes 52447 crash cases in which 134, 3314 and 48999 cases are related to fatalities, injuries and property loss crashes, respectively. These statistics cover crashes occurred in urban highways of Tehran.

## IV. VARIABLES IDENTIFICATION

In order to investigate and model crash statistics some changes in databank structure via access software were made and another databank was prepared in which available data were in the form of nominal or categorical variables that indicates the presence or absence of assumed state. Thus, it is necessary to identify the independent and dependent variables that can be used in the model. The assumed independent variables in these models include 11 types of variables and actually show the crash type of vehicle.

The dependent variable in these models is crash severity and since the fatality crashes are low in numbers, so in the present study, the crash severity categorized into two levels of fatality-injury crashes and the property loss ones. In these models, $Z_1$ and $Z_2$ represent fatality and property damage crashes, respectively. For non-continuous ordered variables, proper models such as Logit, Probit, Loglianer are applied [11], [12]. Table I describes the values of $Z_1$ and $Z_2$.

TABLE I
UNITS FOR MAGNETIC PROPERTIES

| Variable symbol | Variable and its description |
|---|---|
| *Z1* | *Fatality and injury 1, otherwise 0* |
| *Z2* | *Property damage 1, otherwise 0* |

Logit models are useful for conditions where the objective is to predict the occurrence or non-occurrence of a variable regarding other variables associated with it. This model is similar to linear regression models but suitable for the cases where the associated variables are discrete ones. In the current study, two groups of crash severity variables including fatalities and property damages were applied.

In the current study, SPSS[1] software was used and some criteria were selected to obtain the proper model for evaluation. These criteria are as following:

1) $\rho^2$ which as approaches to 1 indicates the best model processing

2) $Sig$[2] which indicates the meaningful level of the obtained coefficient for model variables. In fact, this variable ensures the confidence level that the coefficient of the assumed variable would not be zero. Generally, variables with $Sig$ value of 0.05, i.e. with confidence interval (CI) of 0.95 are statistically meaningful and accepted in the model.

3) Percentage correct which is used in comparing the model predicted values with observations which indicates that what percentage of observations is correctly estimated.

The main objective of selecting the best model is the optimization of criterion by selecting the best conformity with the least complication. In fact, it means the most conformity between the results of the model and the values obtained from observations.

## V. MODELS PRESENTATION AND VALIDATION

To develop the model, at first the correlation of analysis performed among model parameters in order that the similar variables not be used in one model, simultaneously. Then, the modeling process conducted using variables of volume, speed, lane width, sex and age of driver at fault, type of impact collision, how the impact occurred, type of faulty vehicle, lighting, definite cause of crash and variables associated with the condition of lane at the crash site using 2005-2007 data.

It must be mentioned that in order to control the variables entering or exiting the Logit equation, SPSS provides different methods. So, in this study the backward regression method is used. In this method, at first the model includes all independent variables. Then, in every step, one of the variables which cause the least change in $R^2$ value, leaves the model. In this case, the theory must not be denied with proper selection of $R^2$. The true value of change equals to

[1] Statistical package for social science

[2] Significance level

zero (this is done in a predetermined meaningful level in which its pre selected value is $\geq 0.1$ ). Excluding variables through model ceases when any of the excluded residual variables of the model causes a meaningful change in $R^2$. Regarding these explanations and selecting the Backward regression method, the first model (model 1), i,e $LL(\beta)$ (low compared to $LL(c)$), variable meaningfulness of variables except $E_1$ ( $Sig = 0.115$ ) obtained with 95% confidence ( $Sig \leq 0.05$ ) and high correct percentage. Obtained variables of $(\beta)$ were related to fatality probability function ($P_{Khj}$) in which the negative coefficient variables increased the crash severity (fatality probability) and positive coefficient variables decreased the crash severity. The model calibrated with 23 variables after 4 times iteration. Table II defines the used parameters in developing the models.

TABLE II
DEFINATION OF PARAMETERS

| Symbol | Variable |
|---|---|
| V1 | Traffic volume of $\prec 2000 \; vehicle/hour$ |
| S3 | Speed of 40-60 $km/h$ |
| LW3 | Lane width of 15-18 m |
| LW4 | Lane width of 18-22 m |
| M | Male driver |
| AGE1 | Driver age $\prec 25$ years |
| A2 | Type of crash, multi-vehicle |
| B1 | Type of crash, frontal |
| B3 | Type of crash, frontal-lateral |
| VEH2 | Type of vehicle at fault, passenger car, hired car |
| VEH6 | Type of vehicle at fault, truck |
| VEH8 | Type of vehicle at fault, motorcycle, bicycle |
| LIGHT 2 | Lane darkness |
| C1 | Ignoring the length space |
| C2 | Ignoring the width space |
| C6 | Inability to control the vehicle |
| C7 | Violating the confidence speed |
| C8 | Leftward deviation |
| C10 | Going on with rear gear |
| C11 | Technical defect in the vehicle |
| E1 | Lane condition at the place of crash, soil and gravel |
| E7 | Bridge in the crash site |
| E8 | Square in the crash site |

Equations (1) and (2) show the model 1 which has been formed with the obtained coefficients.

$$P_{Khj} = \frac{1}{1 + e^{\Delta U}} \tag{1}$$

$$\begin{aligned}
\Delta U =\ & 3.913 - 0.419V_1 + 0.494S_3 + 0.303Lw_3 \\
& + 0.634Lw_4 - 0.532M - 0.484Age_1 - 0.897A_2 \\
& - 0.549B_1 - 0.266B_3 + 0.505Veh_2 + 0.313Veh_6 \\
& - 3.12Veh_8 - 0.145Light_2 + 0.808C_1 + 0.781C_2 \\
& - 1.719C_6 - 1.849C_7 - 0.722C_8 - 0.637C_{10} \\
& - 0.927C_{11} + 0.334E_1 - 0.683E_7 - 0.681E_8
\end{aligned} \tag{2}$$

For this model, percentage correct and Goodness of fit coefficient have been 94.5% and 89.2%, respectively. In addition to above mentioned criteria, the model has been evaluated with data of 2008 in which the results confirm that the model is proper.

Fig. 1 and 2 demonstrate the distribution of fatality crash probability for the categories with the same characteristics in two states of current situation data and data derived from model, respectively. These figures also show that the frequency of most crash categories is lower than 200 samples and the probability of fatality crash is lower than 20% .
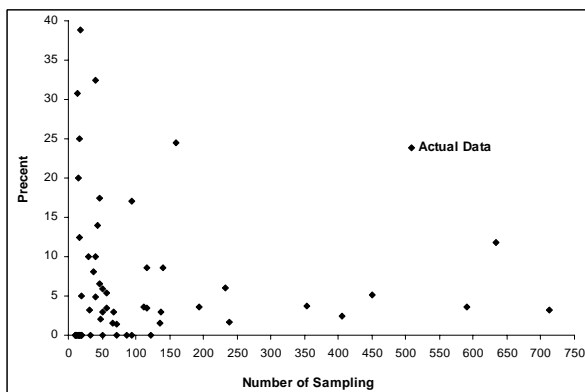


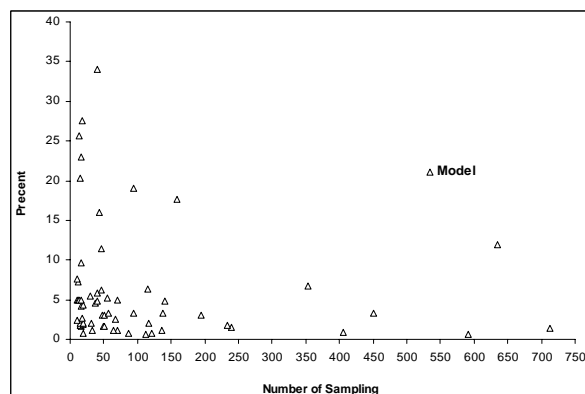Fig. 1 Distribution of fatality crash probability in current situation



Fig. 2 Distribution of fatality crash probability resulted from model

Fig. 3 represents the distribution of fatality crash probability resulted from model and its comparison with the current situation. In this figure, x and y axles are related to crash category with various characteristics and percentage fatality crash probability, respectively.
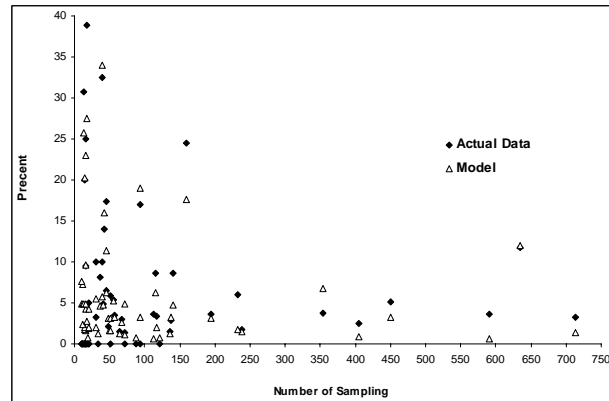


Fig. 3 Distribution of fatality crash probability resulted from model and current situation (magnified)

The distribution of dependent variable of fatality crash probability is demonstrated comparatively in two states of model-derived and current situation proportional with $45°$ line. Generally, it can be concluded that the model is in great conformity with real-life situations and there is an acceptable difference between them.
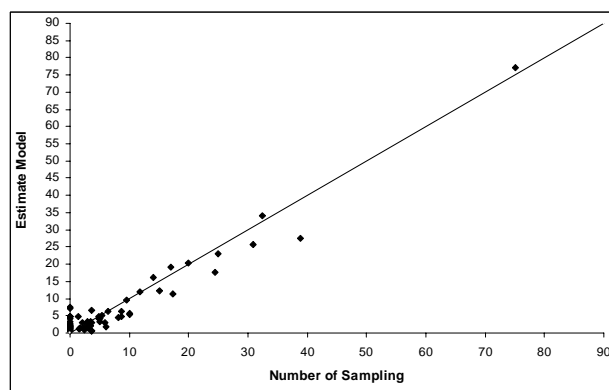


Fig. 4 Comparative figure of fatality crash probability due to the model and current situation

In model 1, $Sig$ of variable $E_1$ is more than the value in model evaluation; however, the obtained model is in acceptable conformity related to real-life. Here, in order to reduce the percentage error in model 1, second model developed by deleting $E_1$ parameter as follows:

Selecting backward regression method, model 2 was developed using the same evaluation rules in model 1. Obtained coefficients ($\beta$) from model 2 relates to the fatality probability function ($P_{Khj}$) in which the negative coefficient variables increase the crash severity (fatality probability) and

positive coefficient variables decrease it. The model calibrated with 22 variables after 4 times iteration. Equations (3) and (4) give model 2 which is formed with obtained coefficients.

$$P_{Khj} = \frac{1}{1 + e^{\Delta U}} \quad (3)$$

$$\begin{aligned}
\Delta U = {}& 3.924 - 0.419V_1 + 0.495S_3 + 0.303Lw_3 \\
& + 0.636Lw_4 - 0.537M - 0.484Age_1 - 0.894A_2 \\
& - 0.549B_1 - 0.267B_3 + 0.503Veh_2 + 0.31Veh_6 \\
& - 3.125Veh_8 - 0.143Light_2 + 0.81C_1 + 0.78C_2 \\
& - 1.719C_6 - 1.845C_7 - 0.722C_8 - 0.637C_{10} \\
& - 0.931C_{11} - 0.69E_7 - 0.688E_8
\end{aligned} \quad (4)$$

For the current model, percentage correct of prediction and Goodness of fit coefficient are 94.9% and 90.6%, respectively. In addition to above criteria, the model was evaluated with 2008 data and the results confirm that the best model is selected. Fig. 5 and 6 show the distribution of fatality crash probability for every category of crashes with similar features in both states of current data and model-derived data, respectively.
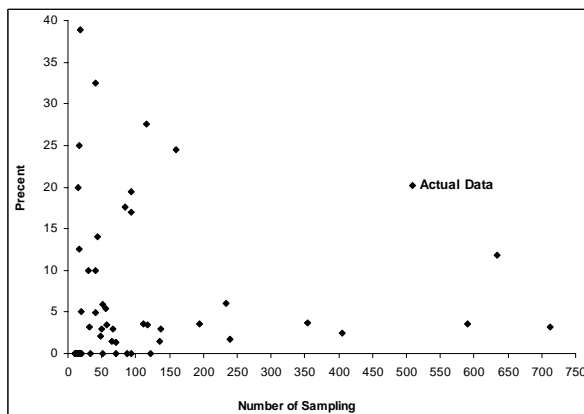


Fig. 5 Distribution of fatality crash probability for current situation
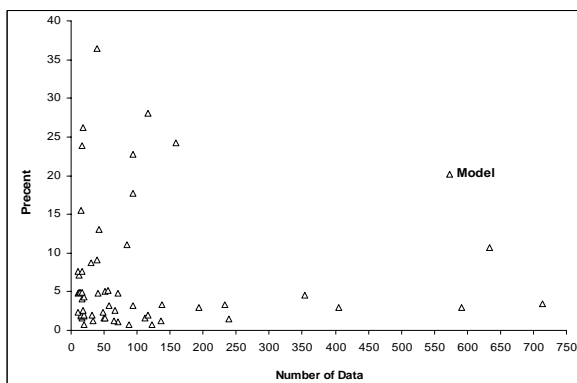


Fig. 6 Distribution of fatality crash probability resulted from model

As depicted in these figures, the frequency of most crash categories is lower than 200 samples and the probability of crash fatalities is lower than 20%. So, Fig. 7 presents the best distribution of fatality crash probability resulted from the model compared with current situation. In these figures, x and y relates to crash category and fatality crash category (in percentage), respectively.
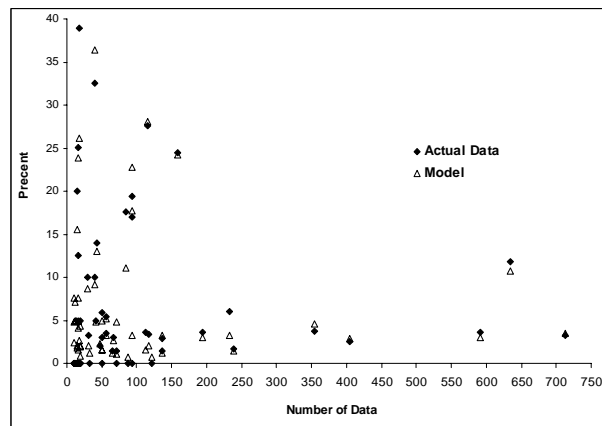


Fig. 7 Distribution of fatality crash probability

Fig. 8 shows the dependent variable of fatality crash probability in both current and model-derived states, comparatively proportional with vertical line $45°$. Generally, regarding the above figures, it can be concluded that the model is in best conformity with real life models and there is an acceptable difference between them.
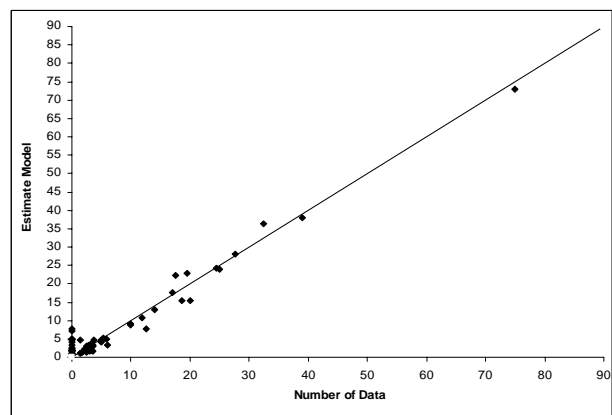


Fig. 8 Comparative figure of fatality crash probability resulted from current and model-derived situations

Comparing two models show that model 2 gives real and accurate results related to model 1. It may be due to omission of $E_1$ variable with $Sig = 0.115$. In fact, it can be concluded that if $Sig = 0$ the model will give more accurate results.

## VI. CONCLUSION

Two models were introduced in this paper that aimed to illustrate the simultaneous influence of variables on the crash severity in urban highways. Developed models indicate that the relationship between crash severity in urban highways and traffic variables same as traffic volume and flow speed, human factors, road, vehicle and weather conditions. Regarding that there is no model in previous researches which can represent the effects of all variables on crash severity of urban highways, the developed models in current study, make it possible to investigate the effects of all the variables on crash severity, simultaneously. So, the models developed can be used to identifying and classify the influential factors in crash severity. Thus, these models suggest that changes in crash severity doesn't occur necessarily by any single dependent parameters, but occur as a simultaneous result of changes in these parameters.

Consider to obtained results from the models it can be concluded that the driver's youth (<25), sex driver, darkness, inability to control the vehicle, leftward deviation, exceeding safe speed, traffic volume of fewer than 2000 vehicles per hour, using rear gear in highways, technical defects of the vehicle, bicycle and motorcycle accidents, accidents on bridges, frontal impact collisions, frontal-lateral crashes, and multi-vehicle crashes can be the variables which increase the crash severity.

Obtained results confirm the effectiveness and capability of the developed models.

## REFERENCES

[1] L. Chang and F. Mannering, "Analysis of Vehicle Occupancy and the Severity Of Rruck and Non-Truck-Involved Accidents," Department of Civil Engineering, 121 More Hall, Box 352700 University of Washington Seattle, Wa 98195, July 17, 1998.
[2] F. F. Saccomanno, S. A. Nassar, and J. H. Shortreed, "Reliability of Statistical Road Accident Injury Severity Models," Transportation Research Record, Issue 1542, pp. 14-23, 1996.
[3] W. Chen and P. P. Jovanis, "Method for Identifying Factors Contributing to Driver-Injury Severity in Traffic Crashes," Transportation Research Record, Issue 1717, pp.1-9, 2000.
[4] K. M. Koekelman and Y. Kweon, "Driver Injury Severityi an Application of Ordered Probit Models," Paper Submitted to Accident Analysis and Preventation, Jan. 2001.
[5] A. Voget and J. Bared, "Accident Models for Two Lane Rural Segments and Intersection," Transportation Research Record, Issue 1635, pp. 18-29, 1999.
[6] K. Kim, L. Nitz and J. L. L. Richardson, "Analyzing The Relationship Between Crash Type and Injuries In Motor Vehicle Collisions In Hawaii," Transportation Research Record, Issue 1467, pp. 9-13, 1994.
[7] A. J. Khttak, P. Kantor and F. M. Council, "Role of Advers Weather In Key Crash Type On Limited: Access Roadways Implications For Advanced Weather Systems," Transportation Research Record, Issue 1621, pp. 15-19, 1999.
[8] H. T. Abdelwahab and M. A. Abdel-Aty, "Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalizes Intersection," Transportation Research Record issue 1746, Paper No.01-2234, pp. 6–13, 2001.
[9] M. A. Abdel-Aty and H. T. Abdelwahab, "Predicting injury severity levels in traffic crashes: a modeling comparison," J. Transp. Eng. vol. 130, no. 2, pp. 204–210, 2004.
[10] D. Delen, R. Sharda and M. Bessonov, "Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks," Accident Analysis and Prevention, vol. 38, pp. 434–444, 2006.
[11] N. Ivan, E. P. Garder and S. Z. Sylvia, "Finding Strategies To Improve Pedestrian Safety in Rural Areas," University of Connecticut and University of Maine, 2001.
[12] Y. J. Kweon and K. M. Kochelman, "The Safety Effects of Speed Limit Changes: Use of Panel Models, Including Speed, Use and Design Variables," The 84th Annual Meeting of Transportation Research Board, Washington D. C., 2005.
[13] Z. Sawallha and T. Sayed, "Statistical Issues in Traffic Accident Modeling," TRB 2003 Annual Meeting, 2003.