IJTE, Vol. 9, No. 3, Winter 2022, pp. 693-711

Research Paper

# Investigating the Effect of Marginal Areas around the Cities on Rural Road Accidents in Iran Using Linear and Logistic Regression Approaches

Hamid Shamanian Esfahani [1], Shahriar Afandizadeh [2,*], Ali Naderan [3]

*Received: 2021/04/30* *Accepted: 2021/10/18*


## Abstract

According to previous studies, 60% to 70% of the total rural road accidents would occur at the city entrance zones in Iran. Therefore, the characteristics of these zones could be considered as effective parameters in rural road accidents. In all prior studies, a 30-km buffer of the cities' entrances has been assumed as the border of the entrance zone. The 30-km buffer could not be considered as the boundary of the influenced area (BIA) of the cites´ entrance for all types of the roads and cities, merely based on aggregate rural road accidents´ data and a traditional definition of the city entrance zone. Determining the BIA for various rural roads with different characteristics using the modelling approach is the innovative aspect of this research. Furthermore, according to their specifications, implementing safety improvements in these areas, not only reduce the number of rural road accidents and fatalities, but also prevent the loss of road safety costs due to the economic problems of Iran. Thus, this study aimed to develop linear and logistic regression models to predict the BIA of rural roads in Iran. The results of this study indicated a fit index value of 0.737 for the linear regression model, and 0.379 and 0.346 for the ordered probit (OP) and multinomial logit (ML) models, respectively. The analysis of significant variables at 95% confidence level, revealed that the access points' density, and the length of adjacent land uses are the most significant variables affecting the BIA.


**Keywords:** Boundary of Influence Area (BIA), Linear Regression Model, Logistic Regression Model, Highway Access Point, Highway Adjacent Land use

* Corresponding author. E-mail: zargari@iust.ac.ir

1 Ph.D. Candidate, Department of Civil Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

2 Professor, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.

3 Assistant Professor, Department of Civil Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

# 1. Introduction

According to the latest statistics, Iran has the population of approximately 80 million, and the death rate caused by traffic accidents in this developing country is 20.5 deaths per 100,000 of individuals [WHO, 2018]. The result of economic studies showed that the direct and indirect costs of accidents and their consequent injuries and fatalities are about $3.6 billion annually (calculated based on 2007 prices). This is equivalent to 6.23% of Iranian gross national product (GDP) in the year 2007 [Ehsani-Sohi, Dashtestaninejad, and Khademi, 2019; Elyasi et al., 2017].

On the other hand, according to statistics published by the road maintenance and transportation organization of Iran [RMTO, 2020], the total number of deaths caused by traffic accidents from March 2019 to March 2020 (based on Iranian official calendar) was 16,947. From this amount, around 66% (equal with 11,186) have died in rural road accidents. Figure 1 shows the trend of changes in rural road accidents and fatalities in Iran during the recent years [RMTO, 2020].

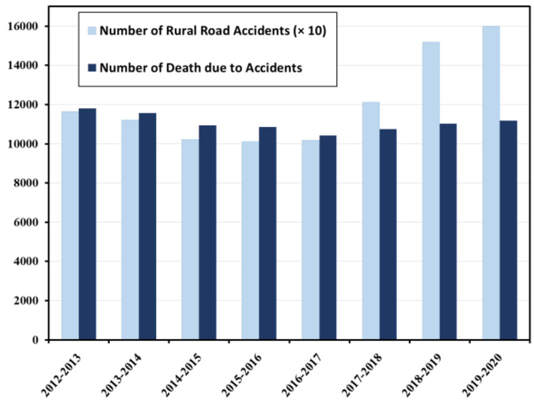



**Figure 1. Changes' tendency in rural road accidents and fatalities in Iran during the recent years (RMTO, 2020)**

A review of previous studies focused on spatial analysis of rural road accidents in Iran, shows that a significant proportion of accidents have been occurred in the marginal areas around the cities [Sajed, Shafabakhsh, and Bagheri, 2019; Elyasi, Saffarzadeh, and Boroujerdian, 2018; Shafabakhsh, Famili, and Akbari, 2016; Effati, Rajabi, and Samadzadegan, 2014; Mohaymany, Shahri, and Mirbagheri, 2013]. In addition, in several road safety studies, it has been mentioned that 60% to 70% of the rural road accidents in Iran have occurred in the cities' surrounding areas, called as the city entrance zone [Ehsani-Sohi, Dashtestaninejad, and Khademi, 2019; Dashtestaninejad, Amiri, and Ehsani-Sohi, 2018; Davoodi and Ahmadi, 2015; Afandizadeh and Golshan-Khavas, 2006; Shafabakhsh and Mousavi, 2006; RMTO, 1999]. In previous studies of the city entrances' safety in Iran, a traditional definition has been used to determine the boundary of the city entrances. Accordingly, the city entrance is the initial 30 km of the rural roads (from the origin city) and the 30 km of the end of the rural roads (from the destination city) [Ahmadi, 2014; Akbarpour, 2013; Afandizadeh and Golshan-Khavas, 2006; Khabiri and Ahmadinejad, 2003]. Therefore, this definition emphasizes a 30-km buffer of the cities´ entrance [Ehsani-Sohi, Dashtestaninejad, and Khademi, 2019]. In all previous studies, a distance of 30 km from the origin or destination city has been considered as the boundary of the cities' entrance area.

To better clarify this issue, the accidents´ cumulative frequency diagram (ACFD) of the Qom-Tehran freeway from March 2017 to March 2018 is shown in Figure 2. The accident data set, which was used for this chart was obtained from the rural traffic police of Iran. Qom-Tehran freeway is a three-lane road placed at each direction with a length of 115 km. Figure 2 demonstrates that 33% of the accidents of this highway have been happened in the first 30 km of the road. Similarly, 35% of these accidents have been occurred in the last 30 km. Consequently, it could be interpreted that 68% of the total accidents of the Qom-Tehran freeway have been occurred within a distance of 30 km from the origin and the destination cities; while, these segments account for 52% of the highway length.

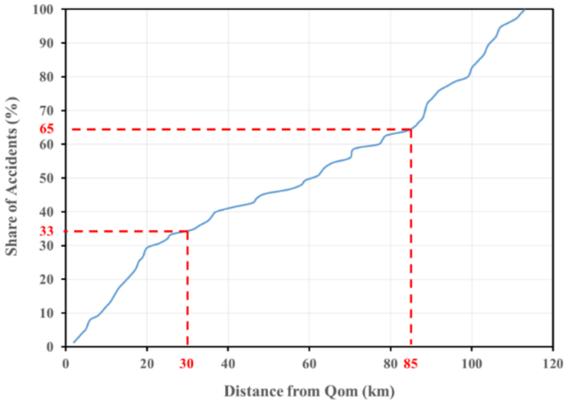


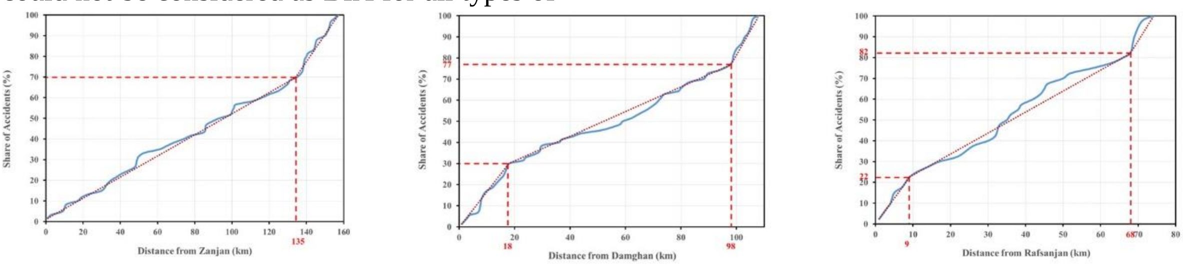

**Figure 2. ACFD of Qom-Tehran freeway for a period of one-year**

In this study, the term ‘marginal area of the city’ has been used instead of the ‘city entrance zone’. According to the preliminary observations, marginal areas around the cities account for a high share of the rural road accidents in Iran; so, they could be regarded as the areas affecting the rural road accidents. To implement the efficient safety improvements in marginal areas and to reduce the accidents and fatalities in surrounding roads, accurate determination of influencing areas on rural road accidents is required.

The definition given in previous studies for the city entrance zone (30-km buffer) is solely based on a preliminary analysis of aggregate rural road accident data. The 30-km boundary could not be considered as BIA for all types of roads and cities. Instead, it could be only measured based on the accident statistics and a traditional definition of the city entrance zone. It can be also investigated in the ACFD of the Qom-Tehran freeway (Figure 2) in which the accident curve is divided into three parts. The first part is the 20-km segment at the beginning of the highway, where 30% of the accidents have occurred. The second part, which includes the distance between 20 and 85 km, is where 35% of the total accidents are occurred. The other 35% of accidents are happened in the third part, which is the last 30 km of this highway.

The slope changing of the curve in the second section compared to the first section, indicates a significant reduction in the accident rate. In the same way, a change in the curve slope is observed in the third section compared to the second section, which indicates another significant increase in the last division. As mentioned above, the BIA is located at the first segment of the road in a distance of 20 km from the Qom city, and the 30-km boundary could not be considered as the BIA for this road's segment.

Other examples of the ACFD rural roads have presented in Figure 3. These diagrams would underline the difference between BIA concept and traditional definition of 30-km kilometers buffer.




a) Freeway of Zanjan-Qazvin (157 km) b) Expressway of Damghan-Semnan (108 km) c) Two-way two-lane road of Rafsanjan-zarand (75 km)

**Figure 3. Examples of ACFD rural roads for a period of one-year**

On the other hand, some variables such as type and density of the adjacent land use, the density of access points and their average distances, road traffic volume, and topography of the region seem to be effective on BIA. Therefore, lack of the accuracy improvement of BIA determination for a variety of roads with different characteristics, and using a constant value for all conditions, could be identified as a gap in previous studies.

Determining the BIA for various rural roads with different characteristics using the modeling approach, which is the innovative part of this research, and implementing safety improvements in these areas based on their specifications, not only reduce the number of rural road accidents and fatalities, but also prevent the loss of road safety costs due the economic problems in Iran. Hence, accentuating the literature gap and considering the importance of BIA scrutiny and its effect on efficiency increase of the road safety improvements in Iran, the main purpose of this study is to provide appropriate models to determine BIA.

## 2. Literature Review

The safety analysis of the city suburbs and residential areas in previous studies can be categorized into two sections. The first part consists of the studies related to safety of the city entrance zones. Their main objectives are investigating the factors affecting rural road accidents at the cities' entrances. The second part includes the studies conducted to identify accident-prone points in the rural roads, in which their results are used to examine the distribution of accident-prone points in the cities' marginal areas.

Khabiri and Ahmadinejad [Khabiri and Ahmadinejad, 2003] in their research determined some factors such as: illegal constructions, various land-use types around the roads, the existence of several access points to the road, lack of complete separation of urban and rural road traffic, different vehicle speeds, and pedestrian traffic, as the reasons of accidents' high rate within 30 km of the city entrances. Shafabakhsh and Mousavi [Shafabakhsh and Mousavi, 2006] considered different traffic flows with different destinations as one of the focal explanations for the high rate of accidents at the city entrances. In another study, Akbarpour [Akbarpour, 2013] analyzed the factors affecting occurrence of the rural road accidents at the cities' entrance in Khuzestan province of Iran by examining the data of Ahwaz-Andimeshk road accidents over three years. It was found that nearly 57% of the accidents on this highway have been occurred at the entrance of Ahwaz and Andimeshk (30 km buffer from the entrance of each city). According to this study, some of the reasons of high accident rate at the entrance zones of these cities, are developing industrial centers on both sides of the roads, high access rates to agricultural farms and residential areas, non-standard intersections, lack of signs and safety equipment, and inadequate asphalt pavement. In another study carried out by Akbarpour, Amini, and Najafi-Alamdarlou [Akbarpour, Amini, and Najafi-Alamdarlou, 2021], some of these parameters were mentioned as effective factors in rural road accidents' occurrence at the cities' entrance as well. Some of the city entrances safety' studies have focused on modeling the number and severity of the accidents in these areas. Afandizadeh and Golshan-Khavas [Afandizadeh and Golshan-Khavas, 2006] have developed the city entrances' safety model using the data of the rural road accidents and regression methods. Based on the results of their study, the exponential function has the best fit with the data of physical and demographic characteristics and the accidents occurred within 30 km of the cities. In more detail, the highway width, vehicle traffic volume, number of access points to the main route, city population, road slope, and road length are introduced as the independent variables in this model. In a study conducted by

Dashtestaninejad, Amiri, and Ehsani-Sohi [Dashtestaninejad, Amiri, and Ehsani-Sohi, 2018], discrete choice models were used to predict the accidents' severity at the city entrances. They detected variables such as age and education, traffic volume, number of speed violations, collision type, and the number of heavy vehicles involved in the accident as affecting parameters on the severity of accidents within 30 km of the city entrance. Ehsani-Sohei, Dashtestaninejad, and Khademi [Ehsani-Sohei, Dashtestaninejad, and Khademi, 2019] used negative binomial and Poisson regression models to predict the number of accidents at the entrance zone of Tehran (the capital of Iran). A comparison of modeling results in that study showed higher efficiency of negative binomial model compared to Poisson regression model. Significant variables in the model were Average Daily Volume (ADT), the proportion of heavy vehicles, daily average number of speed violations, and the density of access points along the highway.

The purpose of the second section of the research is to identify the accident-prone points in rural roads, in which their findings could be used to explore the distribution of black spots in the marginal areas around the cities. Hosseinlou and Sohrabi [Hosseinlou and Sohrabi, 2009] used the adaptive fuzzy-neural inference system (ANFIS) to identify the accident-prone points of the rural roads. Based on the outcomes of their study, one of the variables used in the proposed model (to determine accident-prone points in rural roads) is a variable that explains the effect of residential areas on accidents. In another inquiry, Mohaymany, Shahri, and Mirbagheri [Mohaymany, Shahri, and Mirbagheri, 2013] used Network Kernel Density Estimation (NKDE) to determine the accident-prone points of the Arak-Khomein road in Iran. According to them, 10 km area around the city of Arak and 15 km area around the city of Khomein have a higher risk of accidents compared to other segments of the road. Effati, Rajabi, and Samadzadegan [Effati, Rajabi, and Samadzadegan, 2014] used a fuzzy-neural approach to generate an identification model for high-risk areas in rural roads. To achieve this end, the data of Kouhin-Lowshan road were used, which is a two-lane road with a length of 62 km. According to that study, one of the effective variables in determining high-risk areas in rural roads is the distance from the cities and high- populated centers. Considering the modeling results, one of the five most dangerous areas is located 4 km away from Kouhin city. Also, the parts located within ten km of Kouhin city and six km of Lowshan city are classified into dangerous group segments. These results indicate a high risk of accidents in marginal area of the cities and suburbs. Shafabakhsh, Famili, and Akbari [Shafabakhsh, Famili, and Akbari, 2016] used Kernel Density Estimation (KDE) and Nearest Neighbor Distance Analysis (NNDA) for spatial analysis of Semnan-Garmsar road accidents in Semnan province (Iran). The results of that study showed that in terms of the number of accidents, the most dangerous segment of the road is located within 30 km of Garmsar city; while, in terms of the severity of accidents and fatalities, the most dangerous segment is placed within 10 km around Semnan. Generally, 10 km areas around both Semnan and Garmsar cities are considered as high-risk segments of the Semnan-Garmsar road. Significant share of the rural road accidents in the margins of origin and destination cities has also been verified in other studies conducted by Sajed, Shafabakhsh, and Bagheri [Sajed, Shafabakhsh, and Bagheri, 2019], Elyasi, Saffarzadeh, and Boroujerdian [Elyasi, Saffarzadeh, and Boroujerdian, 2016], Effati, Rajabi, and Samadzadegan [Effati, Rajabi, and Samadzadegan, 2012], and

Boroujerdian, Saffarzadeh, and Abolhasannejad [Boroujerdian, Saffarzadeh, and Abolhasannejad, 2010].

# 3. Methodology

In this study, the linear and logistic regression models were used to develop the BIA determination model. This section provides explanations of the conducted models and their applications in solving the problem. In addition, the process of models' validation is explained.

## 3.1. Linear Regression Model

The linear regression model is used to determine the BIA. The linear regression refers to the model, which relates to a dependent variable and one or more independent explanatory variables. The output of the linear regression model has a continuous value. General form of a linear regression model with k independent variables ($x_1$, $x_2$, ..., $x_k$) and a dependent variable (y) is assumed as Equation 1:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

Where, $\beta_0$ to $\beta_k$ are the regression coefficients, and $\varepsilon$ is the error term of the model, which reflects the influence of some effective factors on dependent variable that is not considered in the model. To estimate the regression coefficients, the least squares' method is utilized.

Usually, the statistical significance of each variable in the linear regression model is assessed by the t-test; while, the whole model is evaluated by using F-test. The coefficient of determination ($R^2$) is also applied to determine the model's goodness of fit.

## 3.2. Logistic Regression Model

The logistic regression model is considered as a classification algorithm that is used to find the relationship between observed variables and a set of discrete classes. In contrast to the linear regression model, which has a continuous dependent variable, the logistic regression utilizes the logistic sigmoid function to calculate the discrete classes' probability of dependent variable.

Regarding the nature of logistic regression models, these models are appropriate to determine BIA. In this case, instead of considering the BIA as a continuous value, the probability of being within a certain range (first or last segment of the highway) is examined. To determine the BIA for the first segment of a rural road connecting the city A to city B, the initial 40 km segment of the road is divided into eight segments with the length of 5 km. The first category from 0 to 5 km, the second category from 5 to 10 km, and this procedure continues to the eighth category from 35 to 40 km. Therefore, the model output determines which segment is more probable to be considered as the BIA location. For example, if the BIA is more likely to be within a 10 to 15 km segment (third category), the boundary of the influence area around the city A for the intended highway, will be 15 km. Similarly, the BIA can also be specified for the end of the road.

Multinomial logistic regression (MLR), which generalizes logistic regression to multiclass problems, is a semi-parametric classification statistic. MLR uses a set of independent variables of any type (e.g., binary, ordinal, continuous) to predict various outcomes' probabilities of a categorically distributed dependent variable. It utilizes the log odds' ratio and an iterative maximum likelihood method to develop the final model and to predict the group membership. It is based on some assumptions, including (1) independent variable with a single value for each case; (2) relatively low collinearity; and (3) the independence of irrelevant alternatives (IIA). IIA defines the odds of preferring one category to another, without considering the presence of other irrelevant alternatives.

Equation 2 represents the tendency of observation i towards the category k in the Multinomial Logit (ML) model, as follows:

$$T_{ki} = \alpha_k + \beta_k X_{ki} + \varepsilon_{ki} \qquad (2)$$

In which, $\alpha_k$ is a constant parameter for the category k of dependent variable, $\beta_k$ shows a vector of parameters for the category k of dependent variable, $X_{ki}$ represents a vector of explanatory variables affecting dependent variable for observation i at the category k, $\varepsilon_{ki}$ denotes a random error term with a Gumbel distribution; i = 1,…, n where n is the total number of observations used in the model.

Equation 3 calculates the probability of each category. Assume $P_i(k)$ is the probability of observation i in the category k, such that:

$$P_i(k) = \frac{\exp(\alpha_k + \beta_k X_{ki})}{\sum_{\forall k} \exp(\alpha_k + \beta_k X_{ki})} \qquad (3)$$

The ordered discrete choice models (i.e., the ordered probit/logit models: OP/OL), which ignores the IIA assumptions, have been used to analyze the multinomial variables with an ordered nature [Pai and Saleh, 2008].

The Op/OL models consider a latent variable z, as demonstrated in Equation 4 to determine the outcome, as the following:

$$z = \beta X + \varepsilon \qquad (4)$$

Where, X is the vector of explanatory variables for each observation, β is the vector of coefficients, and ε is a random error term with a Gumbel distribution. Then, the dependent variable y is estimated by Equation 5:

$$y = \begin{cases} 1, & \text{if } z \le \gamma_1 \\ k, & \text{if } \gamma_{k-1} < z \le \gamma_k \\ K, & \text{if } z > \gamma_{K-1} \end{cases} \qquad (5)$$

Where, $\gamma = \{\gamma_1 \ldots, \gamma_k \ldots, \gamma_{K-1}\}$ are the threshold values of all dependent variable categories; k = 1 …, K; and K is the highest category.

Given the value of X, the probability of dependent variable being in each category could be determined as Equation 6:

$$\begin{cases} p(y=1) = \Phi(-\beta X) \\ p(y=k) = \Phi(\gamma_{k-1} - \beta X) \\ \qquad\qquad -\Phi(\gamma_{k-2} - \beta X) \\ p(y=K) = 1 - \Phi(\gamma_{K-1} - \beta X) \end{cases} \qquad (6)$$

Where, Φ(u) denotes the cumulative density function of the random error term ε evaluated at u. To evaluate the parameters of the OP/OL models, the maximum likelihood method is utilized. As OP and OL models produce remarkably similar results [O'Donnell and Corner, 1996; Pai and Saleh, 2008], only the estimation results of the OP model have been reported in this study.

It is a key assumption in ordered models that the effects of any explanatory variable on different thresholds are consistent or proportional; hence, this is usually termed as the assumption of proportional odds (APO). Based on this assumption, the explanatory variables have the same effect on the odds without considering the threshold [Asare and Mensah, 2020; Sasidharan and Menéndez, 2014; Wang and Abdel-Aty, 2008].

### 3.3. Models' Validation

After the modeling process, the models are validated using the validation dedicated data. As other studies have also focused on modeling [Saheli and Effati, 2021; Sheikholeslami Bondarabadi, and Asadamraji, 2020], the validation process is performed using 20% of the observations that are not used for the development of models. The mean absolute deviation (MAD) and root mean square error (RMSE) were then computed to validate the models' development using validation dataset. They are represented through the following equations:

$$\mathrm{MAD} = \frac{1}{n}\sum_{t=1}^{n}\left|y_{\mathrm{predicted}} - y_{\mathrm{observed}}\right| \quad (7)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(y_{\mathrm{predicted}} - y_{\mathrm{observed}}\right)^2} \quad (8)$$

Where, $y_{predicted}$ and $y_{observed}$ are the predicted and observed value of BIA, respectively, and n is the number of observations in the validation process.

## 4. Database Analysis

To develop BIA determination models for the rural roads, the data obtained from 26 roads in Iran have been used. Among them, seven roads are freeways, nine are expressways, and the other ten roads are two-lanes. The freeways and expressways have two directions; thus, totally 32 freeway and expressway roads were studied in this research.

Also, there are two cities at both ends of each road that their marginal areas' impact on the road has been investigated. As a result, there are totally 64 observations for the freeways and expressways in this study. Furthermore, there are altogether 20 observations for two-lane roads; since, there is only one route between the origin and destination cities. Accordingly, a total number of 84 observations were considered. While determining the case study, the diversity and extent of the characteristics of the roads and cities at both ends of each road, were taken into account.

The data gathered for selected highways included the following issues: the accident data of each road (data for a period of one year from March 2017 up to March 2018), information about the cities at both ends of each road, type of road, topographic information, traffic volume, speed limit, number of lanes, number of access points, average access distance, conditions of adjacent land-uses, separated information of each existing land-use, and geometric characteristics of the highway. Apart from the accident data obtained from the rural traffic police, other required information was received from the Ministry of Roads and Urban Development as well as the Road Maintenance and Transportation Organization of Iran.

The purpose of developing BIA models is to determine the length of the beginning or ending segment of the road; where, the accident rate is significantly different from the middle segment. To define the value of dependent variable for each observation, an ACFD was plotted for each route. Traffic accident data were used to plot these curves, as shown in Figure 2. ACFD presents the proportion of different road segments in the total number of occurred accidents over one year. Therefore, the BIA was specified for the beginning and the end of the route by this curve and through identifying the break-points of the slope. To determine the exact break-points, each route was divided into two km segments, and the number of each segment's accidents was identified. Then, the difference between the number of accidents in each segment and its upstream segment was calculated and divided by the total number of these road accidents, so that the changing rate in the number of accidents was distinguished for different segments of the highway. This index identifies major changes in the accident rate along the route. Finally, it is possible to identify the segment located at the city margin, which has the greatest change in accident rates. The purpose of examining the crash rate changes in the marginal area is to detect the BIA at the beginning and end of intended route. The BIA is determined by its distance from the beginning or end of the road in kilometer. The methodology of BIA determination is characterized using the following equations:

$$S_n = \frac{A_n}{L_n} \quad (9)$$

$$\Delta S_n = S_n - S_{n-1} \quad (10)$$

$$\text{for } m = 1 \text{ to } N$$
$$\text{if } \left(\Delta S_m = \max_{n=1 \text{ to } N}(\Delta S_n)\right) \quad (11)$$
$$\text{then } m = BIA$$

Where, $A_n$ is the number of accidents in segment n, Ln is the length of segment n, N is the total number of segments, and m is the segment in which the BIA is placed.

Notably, the BIA value will only be greater than zero if the ACFD slope in the city margin is greater than the middle segment. Otherwise, the BIA value is considered to be zero. Likewise, if there is no change in the slope of the middle section in comparison with the initial or final sections, the BIA value would be still zero.

A descriptive analysis of the main continuous and discrete variables of the modeling database is presented in Table 1 and Table 2, respectively. For some of the variables introduced in Table 1, such as the number of access points and the length of adjacent land-uses to the road, a fixed value was not provided for the entire route. To determine the value of these variables, the road segment located at the marginal area of the city was divided into equal parts, and the value of the variable was determined for each part, separately. Therefore, considering the maximum length of a city margin equal to 40 km and dividing the initial and final parts of the studied route into five km segments, for each observation, the amount of mentioned variables were separately determined for the segments one to eight. The nearest five km segment to the origin or destination city is always considered as the first segment, and the farthest segment from the city is considered as the eighth segment. It should be noted that for this group of variables, the minimum and maximum values mentioned in Table 1 correspond to the first segment. The value of these variables is also measured cumulatively. For example, in addition to measuring the number of road access points in each five km segment, the number of road access points at the distances of 5 to 40 km around the city is also specified.

Another variable attributed to some of the road features, such as the number of access points and the length of land-uses is defined as the breakpoint of the cumulative diagram (BCD). As was explained for determining the value of the dependent variable for each observation, a cumulative curve was plotted for each feature. These curves illustrate changes in the number of access points and the length of adjacent land-uses in different parts of the road. To determine the break-point of these curves, the same method described for the accident curve was used, and the position of the break-point was identified according to its distance from the beginning or the end of the road in kilometers.

**Table 1. Descriptive analysis of continuous variables of the modeling database**

| Variable | Description | Mean | Min. | Max. | St.d |
|---|---|---|---|---|---|
| BIA | BIA for each observation (in km) | 6.655 | 0 | 30 | 7.895 |
| P | the population of origin or destination city (thousands) | 588 | 1 | 8694 | 1372 |
| A | the area of origin or destination city ($km^2$) | 49 | 0.3 | 645.0 | 102 |
| S | speed limit of the road (km/h) | 109 | 90 | 120 | 11 |
| L | number of lanes in one direction | 2 | 1 | 4 | 1 |
| V | average traffic volume in one lane (veh/hr/lane) | 212 | 31 | 785 | 139 |
| NAP (in segment #) | number of access points in the 5-km segment | 4 | 0 | 9 | 3 |
| NAP-BCD | break-point of access points cumulative frequency diagram (in km) | 9.9 | 0 | 28 | 6.8 |

| Variable | Description | Mean | Min. | Max. | St.d |
|---|---|---|---|---|---|
| DAP (In segment #) | the average distance between access points in the 5-km segment (m) | 2002 | 140 | 5000 | 1738 |
| LALU (in segment #) | length of roads' adjacent land use in the 5-km segment (km) | 2.2 | 0.0 | 5.0 | 2.2 |
| LALU- BCD | break-point of cumulative land use' length diagram (in km) | 9.4 | 0.0 | 38.0 | 9.7 |
| LALU-R (in segment #) | length of roads' adjacent residential land use in the 5-km segment (km) | 0.5 | 0.0 | 4.9 | 0.9 |
| LALU-I (in segment #) | length of roads' adjacent industrial land use in the 5-km segment (km) | 0.4 | 0.0 | 2.5 | 0.6 |
| LALU-A (in segment #) | length of roads' adjacent agricultural land use in the 5-km segment (km) | 1.1 | 0.0 | 5.0 | 1.9 |
| LALU-C (in segment #) | length of roads' adjacent commercial land use in the 5-km segment (km) | 0.2 | 0.0 | 3.0 | 0.5 |
| LHC (in segment #) | length of Horizontal curves in the 5-km segment v | 0.5 | 0.0 | 4.5 | 1.1 |

**Table 2. Descriptive analysis of discrete variables of the modeling database**

| Variable | Description | | Frequency | Percent (%) |
|---|---|---|---|---|
| FI | freeway indicator | freeway = 1 | 28 | 33 |
| | | otherwise = 0 | 56 | 67 |
| EI | expressway indicator | expressway = 1 | 36 | 43 |
| | | otherwise = 0 | 48 | 57 |
| TI | two-lane indicator | two-lane = 1 | 20 | 24 |
| | | otherwise = 0 | 64 | 76 |
| LTI | level terrain indicator | level terrain = 1 | 71 | 84 |
| | | otherwise = 0 | 13 | 16 |
| RTI | rolling terrain indicator | rolling terrain = 1 | 9 | 11 |
| | | otherwise = 0 | 75 | 89 |
| MTI | mountain terrain indicator | mountain terrain = 1 | 4 | 5 |
| | | otherwise = 0 | 80 | 95 |

# 5. Results

## 5.1. Linear Regression Model

Concerning the correlations' analysis results based on the Pearson coefficient and the Spearman coefficient, there is strong positive correlation between BIA and some independent variables such as the city population (P), city area (A), average traffic volume (V), the break-point of access points cumulative frequency diagram (NAP-BCD), and the break-point of cumulative land use' length diagram (LALU-BCD). However, some of these independent variables had a high correlation degree with each other. For example, the correlation coefficient between the city population and city area was significant. Using all variables introduced in Table 1, analyzing the correlation between the variables, and considering other model evaluation parameters, the best linear regression model was obtained for estimating BIA. The best linear regression model is presented in Table 3.

**Table 3. The best calibrated linear regression model**

| Variable | Coefficient | St.d | t-ratio | p-value |
|---|---|---|---|---|
| V (veh/hr/lane) | 0.008 | 0.003 | 2.408 | 0.018 |
| NAP-BCD (km) | 0.405 | 0.067 | 6.019 | 0.000 |
| LALU- BCD (km) | 0.545 | 0.048 | 11.445 | 0.000 |
| Constant | -4.194 | 0.991 | -4.231 | 0.000 |

$$\mathrm{BIA} = -4.194 + (0.008 \times \mathrm{V}) + (0.405 \times \mathrm{NAP} - \mathrm{BCD}) + (0.545 \times \mathrm{LALU} - \mathrm{BCD})$$

The model provides the goodness of fit (Adjusted R2) at 0.737. Based on the results of the significance analysis, all independent variables of the model were significant at 95% confidence level (t ratio at 95% confidence level was 1.96). The result of the F test (78.486) also confirmed the model. The observation-estimation diagram for the linear regression model is presented in Figure 4. This graph was provided based on the values of the studied roads' model and its comparison with real observations. It should be underlined that in this diagram each point can represent one or more specific observations.

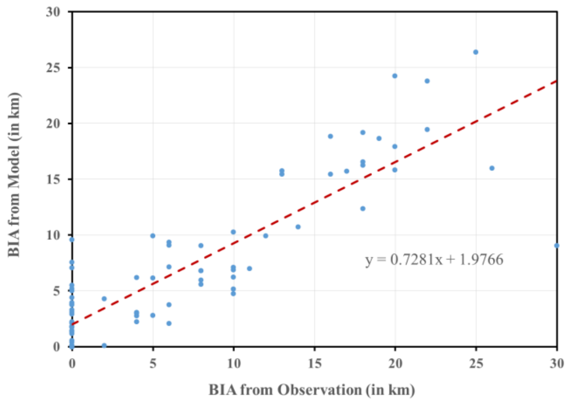



**Figure 4. The observation-estimation diagram for the linear regression model**

## 5.2. OP Model

To develop logistic regression models, the marginal areas of the origin and destination cities for each route were divided into eight 5-kilometer segments. The output of the model revealed the segment where the BIA is more likely to be. Due to the ordered nature of the alternatives (0 to 5 km, 5 to 10 km ..., and 35 to 40 km), the ordered models were more compatible with this case study than other types of logistic regression models. Because of the approximate similarity of the OL and OP results, only the results of the OP model were presented.

For the ordered model estimation, all the variables provided in Table 1 were initially included in the model. Similar to other studies performed on logistic regression modeling [Ghasedi, Sarfjoo, and Bargegol, 2021], the backward approach was used for the process in the way that only those, which were significant at the level of 0.95 were included in the model. Accordingly, the best OP model was obtained as shown in Table 4. In this model, the type of independent variables is continuous. The signs and values of the estimated coefficients of the variables from all models were acceptable. The model fit index value (Adjusted ρ2) was equal to 0.379. The observation-estimation diagram for the OP model is presented in Figure 5. It should be mentioned that in this diagram, each point could represent one or more specific observations.

**Table 4. The best-calibrated OP model**

| Variable | | Coefficient | St.d | t-ratio | p-value |
|---|---|---|---|---|---|
| NAP-BCD (km) | | 0.136 | 0.026 | 5.231 | 0.000 |
| LALU- BCD (km) | | 0.142 | 0.020 | 7.117 | 0.000 |
| Threshold Parameters | $\gamma_1$ | 2.532 | 0.393 | 6.443 | 0.000 |
| | $\gamma_2$ | 3.392 | 0.450 | 7.538 | 0.000 |
| | $\gamma_3$ | 4.672 | 0.605 | 7.722 | 0.000 |
| | $\gamma_4$ | 5.909 | 0.743 | 7.953 | 0.000 |
| | $\gamma_5$ | 6.752 | 0.827 | 8.164 | 0.000 |
| | $\gamma_6$ | 6.931 | 0.847 | 8.183 | 0.000 |
| | $\gamma_7$ | 7.195 | 0.882 | 8.158 | 0.000 |

-2 Log-likelihood at zero = 240.363

-2 Log-likelihood at convergence = 148.484

Adjusted $\rho^2 = 0.379$

$$z = (0.136 \times \mathrm{NAP} - \mathrm{BCD}) + (0.142 \times \mathrm{LALU} - \mathrm{BCD})$$

$$y = \begin{cases} 1, & \text{if } z \leq 2.532 \\ 2, & \text{if } 2.532 < z \leq 3.392 \\ 3, & \text{if } 3.392 < z \leq 4.672 \\ 4, & \text{if } 4.672 < z \leq 5.909 \\ 5, & \text{if } 5.909 < z \leq 6.752 \\ 6, & \text{if } 6.752 < z \leq 6.931 \\ 7, & \text{if } 6.931 < z \leq 7.195 \\ 8, & \text{if } z > 7.195 \end{cases}$$

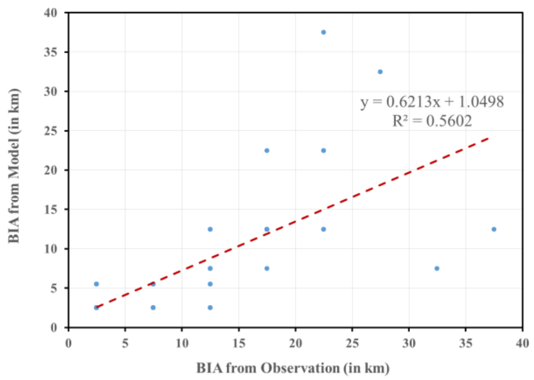



**Figure 5. The observation-estimation diagram for OP model**

As mentioned above, APO assumes that the explanatory variables have the same effect on the odds regardless of the threshold in ordered models. It is necessary to evaluate this assumption in order to apply the OP model properly. Accordingly, the results showed that APO is not held in this dataset; hence, the assumption was rejected. However, APO is frequently rejected particularly when there is a continuous explanatory variable in the model [Allison, 1999].

### 5.3. ML Model

Due to the APO violation in the OP model, the ML method was used to develop the logistic regression model. According to Table 5, after analyzing the significance of the variables at 95% confidence level and considering not correlated independent variables as well as other evaluation criteria of the logistic regression models, the best ML model was obtained The type of independent variables of this model was defined as binary variables. For the eight-level outcome, category eight (35-40) was used as the baseline. Concerning the nature of the problem in this study, the coefficients of explanatory variables in the utility functions were set to be equal. In each category, the explanatory variable's value in the utility function was different from other categories; therefore, the coefficients were equal.

**Table 5. The best-calibrated ML model**

| Variable | Category 1 (0-5) | Category 2 (5-10) | Category 3 (10-15) | Category 4 (15-20) | Category 5 (20-25) | Category 6 (25-30) | Category 7 (30-35) |
|---|---|---|---|---|---|---|---|
| Constant | 30.446 (0.044)[a] | 28.882 (0.023) | 28.976 (0.032) | 28.372 (0.019) | 28.934 (0.027) | 27.773 (0.008) | 27.988 (0.011) |
| NAP-BCD [b] | 2.396 (6.053) | 2.396 (6.053) | 2.396 (6.053) | 2.396 (6.053) | 2.396 (6.053) | 2.396 (6.053) | 2.396 (6.053) |
| LALU- BCD [c] | 2.340 (6.468) | 2.340 (6.468) | 2.340 (6.468) | 2.340 (6.468) | 2.340 (6.468) | 2.340 (6.468) | 2.340 (6.468) |

-2 Log-likelihood at zero = 119.894

-2 Log-likelihood at convergence = 77.203

Adjusted $\rho^2$ = 0.346

$$
T_k = \begin{cases}
30.446 + (2.396 \times \mathrm{NAP-BCD}) + (2.340 \times \mathrm{LALU-BCD}), & \text{for } k = 1 \\
28.882 + (2.396 \times \mathrm{NAP-BCD}) + (2.340 \times \mathrm{LALU-BCD}), & \text{for } k = 2 \\
28.976 + (2.396 \times \mathrm{NAP-BCD}) + (2.340 \times \mathrm{LALU-BCD}), & \text{for } k = 3 \\
28.372 + (2.396 \times \mathrm{NAP-BCD}) + (2.340 \times \mathrm{LALU-BCD}), & \text{for } k = 4 \\
28.934 + (2.396 \times \mathrm{NAP-BCD}) + (2.340 \times \mathrm{LALU-BCD}), & \text{for } k = 5 \\
27.773 + (2.396 \times \mathrm{NAP-BCD}) + (2.340 \times \mathrm{LALU-BCD}), & \text{for } k = 6 \\
27.988 + (2.396 \times \mathrm{NAP-BCD}) + (2.340 \times \mathrm{LALU-BCD}), & \text{for } k = 7
\end{cases}
$$

[a] Values in parentheses are the t-ratio of each estimated parameter.

[b] NAP-BCD is in category # = 1; otherwise = 0.

[c] LALU-BCD is in category # = 1; otherwise = 0.

[d] for the eight outcome levels, category eight (35-40) was used as the baseline and its utility function was equal to zero.

As reported in the table, the value of the model fit index (Adjusted ρ2) is equal to 0.346. The signs and values for estimated coefficients of the variables of all models were considered reasonable. The Small-Hsiao IIA test [Washington, Karlaftis, and Mannering, 2011] was conducted, and it was uncovered that the multinomial logit model cannot be refuted. In addition, it was indicated that the IIA assumption among eight categories could not be rejected at the 0.10 significance level.

Table 6 was used to evaluate the accuracy of the ML model. The rows of this table show different cases; where, the values of the NAP-BCD and LALU-BCD indicator are equal with one. Moreover, the number of observations related to each case of the total available observations was specified. The last column shows the probability of locating BIA in each category based on the real observations and models. The numbers out of the parentheses indicate the values obtained by the real observations, and the numbers in parentheses pinpoint the results of the model. For example, according to the sixth row, in six of 84 observations, the BCD access point is in the range of five to ten km, and the length of BCD adjacent land-use is in the range of zero to five km. In five of these six observations, the BIA was in the range of zero to five km and one case was in the range of five to ten km. On one hand, based on the real observations, the probability of BIA being in the first interval was 83.3%, and the probability of being in the second interval was 16.7%. On the other hand, based on the results of the model, for this specific case, the probability of BIA being in the first interval (zero to five km) was 77.4%, and the probability of being in the second interval (five to ten km) was equal with 17.1%. Similarly, the results of the model can be compared with the real observations in other cases.

**Table 6. The evaluation of ML model accuracy**

| Row | NAP-BCD indicator = 1 | LALU- BCD indicator = 1 | Number of Observations | Probability for each Category (%) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 0-5 | 5-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 |
| 1 | 0-5 | 0-5 | 10 | 100.0 (99.2) | 0.0 (0.2) | 0.0 (0.2) | 0.0 (0.1) | 0.0 (0.2) | 0.0 (0.1) | 0.0 (0.0) |
| 2 | 0-5 | 5-10 | 7 | 85.7 (79.1) | 14.3 (15.7) | 0.0 (1.7) | 0.0 (0.9) | 0.0 (1.6) | 0.0 (0.5) | 0.0 (0.5) |
| 3 | 0-5 | 10-15 | 1 | 100.0 (78.0) | 0.0 (1.5) | 0.0 (17.0) | 0.0 (0.9) | 0.0 (1.6) | 0.0 (0.5) | 0.0 (0.5) |
| 4 | 0-5 | 15-20 | 1 | 100.0 (83.8) | 0.0 (1.6) | 0.0 (1.7) | 0.0 (10.0) | 0.0 (1.7) | 0.0 (0.5) | 0.0 (0.7) |
| 5 | 0-5 | 20-25 | 1 | 100.0 (78.5) | 0.0 (1.5) | 0.0 (1.6) | 0.0 (0.9) | 0.0 (16.4) | 0.0 (0.5) | 0.0 (0.6) |
| 6 | 5-10 | 0-5 | 6 | 83.3 (77.4) | 16.7 (17.1) | 0.0 (1.7) | 0.0 (0.9) | 0.0 (1.7) | 0.0 (0.5) | 0.0 (0.7) |
| 7 | 5-10 | 5-10 | 3 | 66.7 (3.9) | 33.3 (93.2) | 0.0 (0.9) | 0.0 (0.5) | 0.0 (0.9) | 0.0 (0.3) | 0.0 (0.3) |
| 8 | 5-10 | 10-15 | 5 | 0.0 (16.2) | 60.0 (37.1) | 40.0 (38.6) | 0.0 (2.0) | 0.0 (3.6) | 0.0 (1.1) | 0.0 (1.4) |
| 9 | 5-10 | 15-20 | 3 | 0.0 (19.2) | 66.7 (44.1) | 33.3 (4.4) | 0.0 (25.1) | 0.0 (4.2) | 0.0 (1.3) | 0.0 (1.7) |
| 10 | 10-15 | 0-5 | 10 | 80.0 (76.2) | 0.0 (1.5) | 20.0 (18.6) | 0.0 (1.0) | 0.0 (1.6) | 0.0 (0.5) | 0.0 (0.6) |
| 11 | 10-15 | 5-10 | 4 | 0.0 (16.1) | 50.0 (35.1) | 50.0 (40.7) | 0.0 (2.0) | 0.0 (3.6) | 0.0 (1.1) | 0.0 (1.4) |
| 12 | 10-15 | 15-20 | 3 | 0.0 (18.4) | 0.0 (3.9) | 66.7 (46.6) | 33.3 (24.1) | 0.0 (4.1) | 0.0 (1.3) | 0.0 (1.6) |
| 13 | 10-15 | 20-25 | 1 | 0.0 (15.9) | 0.0 (3.3) | 0.0 (40.1) | 100.0 (2.0) | 0.0 (36.3) | 0.0 (1.1) | 0.0 (1.3) |
| 14 | 10-15 | 25-30 | 3 | 0.0 (20.5) | 0.0 (4.3) | 66.7 (51.7) | 0.0 (2.6) | 0.0 (4.5) | 33.3 (14.7) | 0.0 (1.7) |
| 15 | 10-15 | 30-35 | 1 | 0.0 (19.9) | 0.0 (4.1) | 0.0 (50.1) | 100.0 (2.5) | 0.0 (4.4) | 0.0 (1.4) | 0.0 (17.6) |
| 16 | 15-20 | 0-5 | 6 | 100.0 (82.6) | 0.0 (1.7) | 0.0 (1.8) | 0.0 (11.0) | 0.0 (1.7) | 0.0 (0.5) | 0.0 (0.7) |
| 17 | 15-20 | 5-10 | 5 | 40.0 (19.4) | 60.0 (42.1) | 0.0 (4.5) | 0.0 (26.8) | 0.0 (4.3) | 0.0 (1.3) | 0.0 (1.6) |
| 18 | 15-20 | 15-20 | 1 | 0.0 (6.2) | 0.0 (1.3) | 0.0 (1.4) | 100.0 (88.8) | 0.0 (1.4) | 0.0 (0.4) | 0.0 (0.5) |
| 19 | 15-20 | 20-25 | 4 | 0.0 (19.0) | 0.0 (4.0) | 0.0 (4.4) | 50.0 (26.2) | 50.0 (43.5) | 0.0 (1.3) | 0.0 (1.6) |
| 20 | 15-20 | 25-30 | 4 | 0.0 (26.0) | 0.0 (5.5) | 0.0 (6.0) | 50.0 (35.9) | 50.0 (5.7) | 0.0 (18.7) | 0.0 (2.2) |
| 21 | 15-20 | 35-40 | 1 | 0.0 (31.3) | 0.0 (6.5) | 0.0 (7.2) | 0.0 (43.2) | 100.0 (6.9) | 0.0 (2.2) | 0.0 (2.7) |
| 22 | 20-25 | 0-5 | 2 | 100.0 (76.8) | 0.0 (1.6) | 0.0 (1.7) | 0.0 (0.9) | 0.0 (17.9) | 0.0 (0.5) | 0.0 (0.6) |
| 23 | 20-25 | 25-30 | 1 | 0.0 (20.9) | 0.0 (4.4) | 0.0 (4.8) | 0.0 (2.6) | 0.0 (50.5) | 100.0 (15.0) | 0.0 (1.8) |
| 24 | 25-30 | 0-5 | 1 | 0.0 (86.5) | 0.0 (1.7) | 0.0 (1.9) | 0.0 (1.1) | 0.0 (1.8) | 0.0 (6.3) | 100.0 (0.7) |

## 5.4. Discussion

Generally, there are various land-uses in the marginal areas of Iranian cities such as residential, industrial, agricultural, and commercial areas. Furthermore, non-adjacent land-uses are also connected to the main road by the secondary roads. The number of access points and the length of land-uses adjacent to the road usually decrease along with getting more distant from the cities. Previous studies confirm the relationship between the reduced number of access points and the length of adjacent land-uses with reduced rural road accident rates [Saheli and Effati, 2021; Singh et al., 2020; Haghani, Jalalkamali, and Berangi, 2019; Ehsani-Sohi, Dashtestaninejad, and Khademi, 2019; Singh et al., 2018; Fitzpatrick, Lord, and Park, 2010; Gundogdu, 2010]. According to the results of this study, two factors of access points and the length of adjacent land-use, have the most significant impact on determination of BIA. The effect of these factors was measured using the cumulative break-point curve of the access points along with the cumulative break-point curve of the length of land-uses adjacent to the road. However, in addition to these factors, the variable of average traffic volume, is also known as an effective factor on BIA in the linear regression model.

Regarding the coefficient of the average traffic volume in the linear regression model, 1% increase in traffic volume, results in increasing 0.008 times of the existing volume in BIA. When the average traffic volume of the highways is 212 veh/hr/ln, along with 1% increase in volume, the rate of BIA increase would be equal with 1.696%. By examining the coefficient of BCD access point, it can be claimed that along with elevating the distance of BCD access point from the nearest city by 1%, the BIA increases by 0.405 times from the existing value. Therefore, considering the average of BCD access points (9.9 km), BIA increases on average by 4.01%. Similarly, concerning the average of BCD for the length of adjacent land-use (9.4 km), an increase of 1% in the length of BCD adjacent land-use leads to 5.123% increase in BIA.

In the ML model, the signs of descriptive variables are positive, which means that as the distance of cumulative break-points' curve of access points and adjacent land-use from the city increases, the probability of the break-points to be farther from the city as well as the BIA increases.

# 6. Models' Validation

For the model validation, 20% of the whole data was reserved. The output of the OP model determines the interval in which the BIA places. In the ML model, the most possible category to occur determines the model output. Also, to compare the results of the linear and logistic regression models, the output of the linear regression model was reported as intervals. For example, if the BIA value from the linear regression model for an observation is seven km (i.e. the boundary of the marginal area affecting the road accidents is seven km from the origin or destination city), the BIA of the road is placed in the five to ten km interval. Table 7 shows the values of MAD and RMSE for the linear and logistic regression models. The computed MAD values of observed and predicted BIA intervals were ranged from 0.37 to 0.44. The results were also reported for RMSE (ranging from 0.81 to 0.94). Regarding the interval rate (1-8), the deviation results seem to be reasonable.

From the results, it can be seen that all three models have approximately the same accuracy in predicting the BIA category. However, since the linear regression model can provide the output as a continuous value, it is more efficient in estimating BIA. In addition, due to the APO violation in the OP model, it is recommended to use the ML model as a logistic regression model for prediction of the BIA.

**Table 7. The Models' Validation by MAD and RMSE Measures**

| Model | MAD | RMSE |
|---|---|---|
| Linear Regression | 0.37 | 0.81 |
| Ordered Probit | 0.44 | 0.93 |
| Multinomial Logistic | 0.40 | 0.94 |

# 7. Conclusion

Considering previous studies focused on high proportion of the role of the cities' surrounding areas in the total rural road accidents in Iran, these areas should be regarded as marginal areas affecting the rural road accidents. To do so, determining BIA is of great importance; since, gaining a better understanding of the marginal zones around Iranian cities, and taking appropriate safety measures based on their specifications, not only reduces the rural road accidents and fatalities, but also prevents the costs associated with road safety, particularly having serious economic problems in Iran. Therefore, this study proposed suitable models to predict the value of BIA for the rural roads in Iran. To achieve this end, the accident data (during one year from March 2017 to March 2018), and required information from the rural roads was used. The linear regression method was applied to develop the BIA prediction model. According to the results, the main factors affecting the BIA consist of the number of access points, the length of adjacent land-use, and the average traffic volume. The goodness of fit index of the model was 0.737. In addition, the t-test analysis confirmed the significance of all independent variables with a confidence level of 95%.

Then the logistic regression method was used to develop the BIA prediction model. In this case, instead of considering the BIA as a continuous value, the probability of being within a certain range of first or last segment of the road was defined. For this purpose, the road segments located at the margin of the cities were divided into eight segments of five km. Due to the ordered nature of dependent variable, the OP method was first adopted to develop a logistic regression model. Although, the results of variables' significance analysis in the OP model and the value of the model's fit index (0.379) indicated a high utility of the model; however, because of violation of ordered models' basic assumption, the OP model was rejected, and the ML method was used to develop the logistic regression model. According to the results of the research done on the significance of variables, the value of the fit index (0.346), and the IIA assumption were able to confirm the ML model. Based on the outcomes of the model, the main factors affecting the BIA consisted of the number of access points and the length of adjacent land-use along the road.

The analysis of significant variables at 95% confidence level, revealed that the access points' density, and the length of adjacent land uses are the most significant variables affecting the BIA. These factors lead to different values of BIA for different roads. While previous studies considered 30-km boundary as BIA for all types of roads and cities [Ehsani-Sohi, Dashtestaninejad, and Khademi, 2019; Dashtestaninejad, Amiri, and Ehsani-Sohi, 2018; Davoodi and Ahmadi, 2015; Afandizadeh and Golshan-Khavas, 2006; Shafabakhsh and Mousavi, 2006].

The MAD and RMSE indices were used for validating the models. The results of validation showed that all three models have almost the same accuracy in predicting BIA. However, since the linear regression model is better able to provide the output as a continuous value, it is more efficient in estimating BIA.

This study was the first attempt to define the concept of BIA and to determine it for various types of roads using the modeling approach. Further scientific works applying other modeling methodologies may contribute to a better perception of these areas around the cities. This may help the researchers to compare different outcomes and decide what is best to be taken. Investigating the fatalities due to accidents at the entrances of cities and determining the BIA on different roads with different characteristics based on casualties is another thoughtful issue that could be addressed in future studies.

## 8. References

– Afandizadeh, S. and Golshan-Khavas,

R. (2006) “Developing a model for determination the road safety index in city entrance zones”, Journal of Transportation Research, Vol. 3, No. 4, pp. 223-242. (In Persian)

– Ahmadi, P. (2014) “Prioritization of proposed solutions in order to reduce road accidents on cities' entrance by multi-criteria decision making method”, MSc Thesis, Shomal University, Amol, Iran. (In Persian)

– Akbarpour, M. (2013) “Studies of road accidents on cities' entrance (case study: Ahvaz–Andimeshk road)”, MSc Thesis, Shahid Chamran University, Ahvaz, Iran. (In Persian)

– Akbarpour, M., Amini, A. and Najafi-Alamdarlou, M. (2021) “The review of effective factors in the occurrence of accidents at the entrance of cities (case study: Northwest exit road of Ahvaz)”, The 6th International Conference on Advanced Research in Science Engineering and Technology. (In Persian)

– Allison, P.D. (1999) “Logistic regression using the SAS system: Theory and Application”, Cary NC, SAS Institute Inc.

– Asare, I.O. and Mensah, A.C. (2020) “Crash severity modelling using ordinal logistic regression approach”, International Journal of Injury Control and Safety Promotion, pp. 1-8.

– Boroujerdian, A.M., Saffarzadeh, M. and Abolhasannejad, V. (2010) “Developing a model for prioritising high crash road segments”, Proceedings of the Institution of Civil Engineers-Transport, Vol. 163, No. 1, pp. 19-29.

– Dashtestaninejad, H., Amiri, M. and Ehsani-Sohi, M. (2018) “Discrete choice models to predict severity of accidents near the city of Tehran (30 kilometer buffer)”, The 17th International Conference on Traffic and Transportation Studies, Tehran, Iran. (In Persian)

– Davoodi, S.R. and Ahmadi, P. (2015) “Prioritizing suggested strategies to reduce road accidents at the cities entrance using analytic hierarchy process”, International Journal of Asian Social Science, Vol. 5, No. 12, pp. 706-714.

– Effati, M., Rajabi, M.A. and Samadzadegan, F. (2012) “Developing a novel method for road hazardous segment identification based on fuzzy reasoning and GIS”, Journal of Transportation Technologies, Vol. 2, No. 1, pp. 32-40.

– Effati, M., Rajabi, M.A. and Samadzadegan, F. (2014) “A geospatial neuro-fuzzy approach for identification of hazardous zones in regional transportation corridors”, International Journal of Civil Engineering, Vol. 12, No. 3, pp. 289-303.

– Ehsani-Sohi, M, Dashtestaninejad, H. and Khademi, E. (2019) “Effects of roadway and traffic characteristics on accidents frequency at city entrance zone”, International Journal of Transportation Engineering, Vol. 7, No. 2, pp. 139-152.

– Elyasi, M. R., Saffarzadeh, M. and Boroujerdian, A. M. (2016) “A novel dynamic segmentation model for identification and prioritization of black spots based on the pattern of potential for safety improvement”, Transportation Research Part A, Vol. 91, pp. 346–357.

– Elyasi, M. R., Saffarzadeh, M., Boroujerdian, A. M., Semnarshad, M. and Mazaheri, M. (2017) "Prioritization of suburban accident factors based on analytical network process", International Journal of Transportation Engineering, Vol. 5, No. 2, pp. 197-209.

– Elyasi, M.R., Saffarzadeh, M. and Boroujerdian, A.M. (2018) "Assessing the interrelations of traffic collisions' risk factors", Proceedings of the Institution of Civil Engineers-Transport, Vol. 171, No. 6, pp. 309-318.

– Fitzpatrick, K., Lord, D. and Park, B.J. (2010) "Horizontal curve accident modification factor with consideration of driveway density on rural four-lane highways in Texas", J Transp Eng, Vol. 136, No. 9, pp. 827-835.

– Ghasedi, M., Sarfjoo, M. and Bargegol, I. (2021) "Prediction and analysis of the severity and number of suburban accidents using logit model, factor analysis and machine learning: a case study in a developing country", SN Applied Sciences, Vol. 3, No. 1, pp. 1-16.

– Gundogdu, I.B. (2011) "A new approach for GIS-supported mapping of traffic accidents", Proceedings of the institution of civil engineers-transport, Vol. 164, No. 2, pp. 87-96.

– Haghani, M., Jalalkamali, R. and Berangi, M. (2019) "Assigning crashes to road segments in developing countries", Proceedings of the Institution of Civil Engineers-Transport, Vol. 172, No. 5, pp. 299-307.

– Hosseinlou, M.H. and Sohrabi, M. (2009) "Predicting and identifying traffic hot spots applying neuro-fuzzy systems in intercity roads", International Journal of Environmental Science & Technology, Vol. 6, No. 2, pp. 309-314.

– Khabiri, M. and Ahmadinejad, M. (2003) "Investigating the causes of accidents at the cities' entrance in Iran", The 10th Students Conference on Civil Engineering, Tehran, Iran. (In Persian)

– Mohaymany, A.S., Shahri, M. and Mirbagheri, B. (2013) "GIS-based method for detecting high-crash-risk road segments using network kernel density estimation", Geo-spatial Information Science, Vol. 16, No. 2, pp. 113-119.

– O'donnell, C.J. and Connor, D.H. (1996) "Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice", Accident Analysis & Prevention, Vol. 28, No. 6, pp. 739-753.

– Pai, C.W. and Saleh, W. (2008) "Exploring motorcyclist injury severity in approach-turn collisions at T-junctions: Focusing on the effects of driver's failure to yield and junction control measures", Accident Analysis & Prevention, Vol. 40, No. 2, pp. 479-486.

– RMTO (Iran Road Maintenance and Transportation Organization) (1999) "Analysis of accidents at city entrance zones", Tehran, Iran. (In Persian)

– RMTO (Iran Road Maintenance and Transportation Organization) (2020) “Road Statistical Yearbook”, http://www.rmto.ir/Pages/Home.aspx. (In Persian)

– Saheli, M.V. and Effati, M. (2021) “Segment-Based Count Regression Geospatial Modeling of the Effect of Roadside Land Uses on Pedestrian Crash Frequency in Rural Roads”, International Journal of Intelligent Transportation Systems Research, Vol. 19, No. 2, pp. 347-365.

– Sajed, Y., Shafabakhsh, G. and Bagheri, M. (2019) “Hotspot Location Identification Using Accident Data, Traffic and Geometric Characteristics”, Engineering Journal, Vol. 23, No. 6, pp. 191-207.

– Sasidharan, L. and Menéndez, M. (2014) “Partial proportional odds model-An alternate choice for analyzing pedestrian crash injury severities”, Accident Analysis & Prevention, Vol. 72, pp. 330-340.

– Shafabakhsh, G. and Mousavi, M. (2006) “Investigating the causes of accidents at the cities' entrance and organizing surrounding areas for immunization”, The 13th Students Conference on Civil Engineering, Kerman, Iran. (In Persian)

– Shafabakhsh, G.A., Famili, A. and Akbari, M. (2016) “Spatial analysis of data frequency and severity of rural accidents”, Transportation Letters, pp. 1-8.

– Sheikholeslami, S., Bondarabadi, M.A. and Asadamraji, M. (2020) “A rural road accident probability model based on single-vehicle hazard properties including hazard color and mobility: a driving simulator study”, Journal of advanced transportation.

– Singh, G., Pal, M., Yadav, Y. and Singla, T. (2020) “Deep neural network-based predictive modeling of road accidents”, Neural Computing and Applications, pp. 1-10.

– Singh, G., Sachdeva, S.N. and Pal, M. (2018) “Support vector machine model for prediction of accidents on non-urban sections of highways”, Proceedings of the Institution of Civil Engineers-Transport, Vol. 171, No. 5, pp. 253-263.

– Singh, G., Sachdeva, S.N. and Pal, M. (2018), “Comparison of three parametric and machine learning approaches for modeling accident severity on non-urban sections of Indian highways”, Adv Transp Stud Int J Sect B 45, pp. 123-140.

– Wang, X. and Abdel-Aty, M. (2008) “Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models”, Accident Analysis & Prevention, Vol. 40, No. 5, pp. 1674-1682.

– Washington, S., Karlaftis, M. and Mannering, F. (2011) “Statistical and Econometric Methods for Transportation Data Analysis, 2nd ed”, Chapman and Hall/CRC, Boca Raton,FL.

– WHO (World Health Organization) (2018) “Global Status Report on Road Safety”, https://www.who.int/publications/i/item/global-status-report-on-road-safety-2018.