Logistic Analysis of Accident Severity Models for Intercity Roads

M. IBRAHIM [1], I. M. RAMADAN [2] AND N. M. SALLEH 3

[1] Engineer, M.Sc. in Civil Engineering 2010, Cairo University

[2] Associate Professor, Civil Engineering Department, Faculty of Engineering at Shoubra, Banha University

[3] Professor, Civil Engineering Department, Faculty of Engineering at Shoubra, Banha University

ABSTRACT

The main objective of this study is to model accident severity in Egypt. Accident severity models in Egypt as soon as many developing countries are limited and need more effort. Data collected from Egyptian desert roads for the period from 2012 until 2015. Three different desert roads are selected in this research (Alexandria, Ismailia, and Suez) for data collection. Logistic regression analysis was performed to estimate accident severity model. All road, driver, environmental, and vehicle characteristics were studied to choose the most effective variables in the model. Analysis was done using the application of SPSS computer software Version [22]. Various statistical tests was applied to the estimated model to check its strength. Tests showed that the estimated logistic model for accident severity is statistically accepted.

KEYWORDS: Accident modeling,
Accident Rate,
Severity model,
Egyptian desert roads,
Effective accident variables,
Logistic analysis.

1. INTRODUCTION

Traffic accidents are among the major and most dangerous problems all over the world. It causes loss of life, health damage of properties, economic loss, and psychological effect on victims and their families. Road traffic accident victims and injuries is a worldwide public health problem. Road accidents represent the third largest cause of death worldwide. The World Health Organization (WHO) considers road accidents as one of the world's largest public health problems. Every year the lives of approximately 1.25 million people are cut short as a result of a road traffic crash [1].

Road traffic accidents have been increasing sharply in Egypt specially with increasing number of owned vehicles. Desert road accidents cost the Egyptian economy a lot of losses [2].

Decision makers and highway officials need accurate information about relationships between accidents and its causes. Developing accident models can help in predicting accidents effectively and allow decision makers to provide road safety measures and policies. Accident models are usually developed as severity models or frequency models. Severity models can predict the degree of severity at a given location in addition to the determination of effective factors in accident causes.

Accident severity models in Egypt are very rare or almost not exist. Therefore, the scope of this research is the estimation of accident severity models for the intercity roads in Egypt. Logistic regression analysis is the most communally used method in estimating accident severity models. Accordingly, this research focuses on estimating an accident severity models using Logistic analysis for intercity roads in Egypt.

2. LITERATURE REVIEW

Factors affecting roads traffic safety are drivers, vehicles, roadway sections characteristics, and environmental conditions. In Egypt, these factors are responsible for accidents occurrence by the percentages of 74%, 17% and 9% for driver, vehicle and road characteristics, and environmental factors respectively [3]. In general, human factors are responsible for 65% to 81% of traffic accidents [5].

Behairy, mentioned that the percentage of accidents between twenty and fifty years age is about 75 %, which is relatively a very high value. This age has the great effect on the national productivity in all countries. In addition, Behairy, found that male fatalities equals six times female fatalities, and male injuries equals five times female injuries[6].

Abou EL-Naga, concluded that many faults in the vehicle could cause accidents like turning flasher, quality of vehicle doors locking, tires, brakes, etc. Any defect in one or more of these factors may cause an accident happening [4].

Jenkins, D., concluded that women were less skillful and less able to execute difficult maneuvers, while men were more likely to drive too fast, and overtake improperly [11].

Babkov [7] concluded that the accidents rate increases as the width of the roadway decreases. Noland and Oh [20], concluded as the number of lanes increase the number of fatalities increase. The same conclusion was gotten by Jian Xu et al [21]. The situation will be reversed in case of shoulder width. The number of accident will decrease as the shoulder width increase [10].

Abou EI-Naga, found that as the daily traffic volume (veh/day) increases, accidents rate (AR) decreases [4]. Gwynn, assumed that a U shaped relationship exists between traffic flow and accident rates on four-lane sections [17]. Ceder and Livneh found that the relationships between accident rates and hourly traffic flow are different [18]. Yousry, concluded that an increase of the percentage of trucks and buses will reduce the speed and the level of service on the road [8]. Hany M. Hassan, et al, stated that heavy trucks are usually involved in severe and fatal crashes occurring [19].

Eluru et al., found that darker periods often lead to higher accident severity [22]. Plainis et al., stated that night time increase accident frequency. He also stated that the degree of lighting affect number of fatalities and increase stopping distance as a result of the increasing reaction time [23].

Accident severity is always measured based on its category. For example accident may be classified as fatal, injury, or property damage accident. Consequently, models that can represent accident severity should be suitable to represent this categorization like logit or probit models [9].

Logistic regression techniques are used to model probabilistic systems to predict future events. These models are direct probability models that have no requirements on the distributions of the explanatory variables or predictors [12].

Traditional logit models do not impose the sometimes unrealistic parameter

ISSN 2229-5518

restrictions that traditional ordered probability models do [13, 14]. However, its applicability is limited by the assumption that all error terms in severity functions are independent and identically distributed. This means that the logit model is particularly susceptible to correlation of unobserved effects from one accident severity level to the next; YUE LIU, [14]. This assumption may not be valid in some traffic safety studies. Logistic regression models have been used in various shapes for accident severity modeling. It has been used in the form of multinomial, binary, ordered, ordinal, and nested logit models [12, 13, 14, 15].

The logit model is limited in three important ways. It cannot represent random taste variation. It exhibits restrictive substitution patterns. In addition, it cannot be used with panel data when unobserved factors are correlated over time for each decision-maker. In contrast, Probit models solve all the three issues [12, 13, 14, 15].

Probit models can handle random taste variation, they allow any pattern of substitution, and they are applicable to panel data with temporally correlated errors. The only limitation of probit models is that they require normal distributions for all unobserved components of utility. Therefore, in some situations, normal distributions are inappropriate and can lead to unsuitable forecasts [16].

The most common approach for the derivation of such models is to start by specifying a hidden variable, Z, which is used as a basis for modeling the ordinal ranking of data. This unobserved variable is most often specified as a linear function for each crash observation, such that $Z=X\beta$, where X is a vector of variables determining the discrete ordering for each crash observation β, is a vector of estimable parameters, and ε is a disturbance term [14].

Both ordered logit and probit models are shared in restrictions coming from their previous assumptions. One of the primary assumptions of both ordered logit and probit models is that the error variances are assumed to be the same for all data. If this assumption is not satisfied the parameter estimates will not be correct and the standard errors are incorrect. To correct this, employing a heterogeneous choice model which reduces the assumption by obviously specifying the determinants of location-scale. Another important assumption share with both logit and probit models is that the relationship between each pair of outcome groups is the same; Chao Wang [9].

In conclusion, logistic analysis is useful for situations in which someone wants to be able to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. Logistic analysis is suited to models where the dependent variable is distributed. Logistic analysis coefficients can be used to estimate odds ratios for each of the independent variables in the model. In addition, is preferred when assessing the contribution of variables because it is less affected by variance, covariance, and inequalities across group. It is able to handle categorical variables easily. Logistic analysis determines the impact of multiple independent variables presented simultaneously to predict membership of one or other of the two dependent variable categories. In what follows, logistic regression model for accidents will be estimated for intercity roads in Egypt.

3. METHDOLOGY

3.1 Modeling variables

The objective of this model is to calibrate accident severity model. It measures the probability of accident occurrence in a given road section. The dependent variable that has been considered in the research is the accident rate. The accident rate was estimated according to the following equation;

$$\textit{Accident rate (AR)} = \frac{\textit{Number of accidents} \times 10^6}{AADT \times 365 \times N \times L} \qquad (3)$$

Where:

AADT : Average Annual Daily Traffic

N : Number of years considered

L : Section length in kilometer

Accordingly, the outcome of the estimated model is the probability of accident occurrence in a given section per vehicle per day per km in millions. This will be termed as P(AR).

This research focuses on severity accident modeling in intercity Egyptian roads. Data was collected from three selected desert roads for accident modeling; Alexandria, Ismailia and Suez. The available data which was used for analysis and modeling are recorded for the period (2012 up to year 2015).

The following variables have been considered in this research to be checked as independent variables:

- Weather conditions (clear – fog – rain).
- Traffic conditions (Average Annual Daily Traffic (AADT) – Heavy Trucks (HT) – Time “ day or night”)

- Human factor (age – gender).
- Geometrical factors (Shoulder Width (SW) – Median Width (MW) – Pavement Width (PW) – Light Condition (LC) – Surface Condition of road (SC) – Lane Marking (LM) – Curves).

Accidents models are calibrated using the application SPSS computer software Version.22 for determining the accident severity model.

3.2 Data Collection

Zegeer mentioned that it is very important to study accident records over a sufficient period of time, in the US, road authorities used time periods for data ranging from one to five years [10].

Therefore, Accident data used in this studying were obtained from recorded data in the “Egyptian General Authority for Roads, Bridges, and Land Transport, also from General Administration of Traffic “for the period between 2012 and 2015 for the three desert roads. These data include accident characteristics, traffic characteristics, and environmental characteristics.

Another type of data which are the geometric characteristics of the accident locations were collected through field surveying. A field survey form has been designed for this purpose. The form and the collected data using are shown in Table (1).

Table 1. Field Survey Collecting Data Form

Road name :	**From km to km**

Factor		**Value**		
1-	Pavement width			
2-	Shoulder width			
3-	Median width			
4-	Road curvature	straight		
		curved		
5-	Lighting	Two Side		
		One Side		
		Not existing		
6-	Lane marking	good		
		bad		
		Not existing		
7-	Surface condition of road	good		
		bad		
		Not existing		

The surveyed data was arranged and classified and entered to SPSS software for the purpose of modeling a logistic regression model in most effective variables.

4. DATA ANALYSIS AND RESULTS

4.1 Determination of Accident Severity model for intercity roads by logistic Analysis

Too many models have been checked on SPSS software for the purpose of getting the most effective logistic regression model. Results of the best model are developed in model (4). It is clear from this model that it includes seven variables. These variables are explained in a previous section.

$$P(AR) = \frac{exp^{(16.286 - 0.1\ PW - 0.316\ MW - 0.0001\ AADT + 0.318\ HT + 0.742SC + 0.504\ LM + 0.294\ LC)}}{1 + exp^{(16.286 - 0.1\ PW - 0.316\ MW - 0.0001\ AADT + 0.318\ HT + 0.742SC + 0.504\ LM + 0.294\ LC)}} \qquad (4)$$

4.2 Testing the model using Wald test

The Wald test (also called the Wald Chi-Squared Test) is a method to find out if explanatory variables in a model are significant. “Significant” means that they add something to the model; variables that add nothing can be deleted without affecting the model in any meaningful way. The test can be used for a multitude of different models including those with binary variables or continuous variables. This test can also be inverted to obtain confidence regions for coefficient value parameters. Table (2) shows Parameter estimates effect of each significant independent variables. If this test statistic has magnitude larger than 1.96, then the coefficient is statistically significant (at the 95% confidence level).

Table 2. Estimated Parameters of independent variables for general model

Variables	B	Wald	Sig.	Exp(B)
Constant	16.286	71.838	0.000	$118.3*10^5$
PW	– 0.100	4.817	0.028	0.905
MW	– 0.316	7.609	0.006	0.729
AADT	– 0.0001	51.091	0.000	1.000
HT	0.318	53.701	0.000	0.728
SC	0.742	22.413	0.000	0.476
LM	0.504	8.613	0.003	1.656
LC	0.294	4.493	0.034	0.745

It is clear from the table that results shows that all explanatory variable are significant. This reflects the influence of the independent variables on the dependent variable. An exponential form of the coefficient Exp (B) (odds ratio) represents the ratio-change in the odds of the accident severity for a one-unit change in the predictor.

4.3 Testing the model fit

The classification results table shows that correctly classified of the total sample volume is 74.7% (707/947). The percentage of correctly classified evaluates the goodness of fit for the model as in table (3). This table shows results derived of Logistic function to classify observations. It contains two classification group functions for the correctly classified data with summation (707) observations used to fit the model. The probability to happen an accident represented by group (Yes) equal (531) observations, with percentage 88.5 % of correctly classified. However, value of no probability to happen an accident represented by group (No) equal (176) observations, with percentage 50.9 % of correctly classified.

Table 3. Classification results of fit for Logit model

Observed		Predicted Group Membership: No	Predicted Group Membership: Yes	Total %
General model	No	176	171	50.9 %
	Yes	69	531	88.5 %
	Percentage correct			74.7 %

The sensitivity of logit model is another test to evaluate the goodness of fit .For measuring the sensitivity of this logit model the area under ROC curve (Receiver Operating Characteristic) determined which equal 80.1 % as shown in the next figure and it's a good value which represent a strong model.

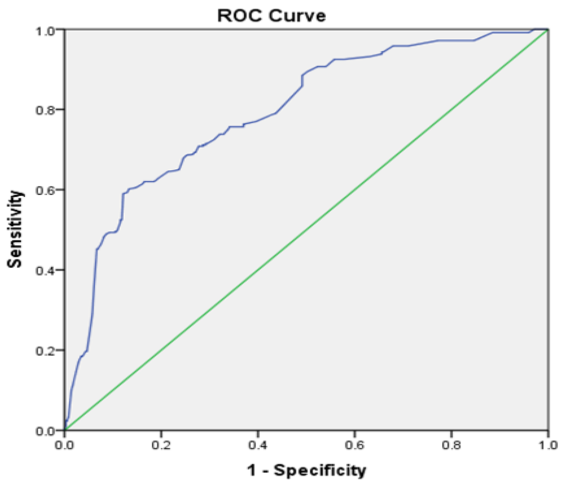

Fig. 1. ROC curve which measuring sensitivity of logit model.

5. CONCLUSIONS

Reference to the above analysis and discussions, one can conclude the following:

- Results showed that the most important factors that affected accident severity at intercity roads in Egypt arranged in a descending order:

  1- LM lane marking
  2- AADT annually average daily traffic
  3- PW pavement width
  4- LC light condition
  5- MW median width
  6- HT percentage of heavy traffic
  7- SC surface condition

- Logistic analysis represents a suitable analysis method for accident severity model.
- It was clearly evident that bad or hidden lane mark is one of the main causes of increasing severity accidents.
- Accident severity increase when lighting condition is absent or poor.
- The current study should be utilized when constructing new intercity roads in Egypt in order to predict accident severity in the design phase. Several geometrical elements could be estimated such as pavement, shoulder and median width, road surface condition, lane marking, and light condition.in order to minimize the probability of accident occurrence.
- Percentage of trucks in the traffic volume increases accident rate.

- The human factor has a major influence on accidents which can be a result of driver negligence of driving rules, un-appreciating landed signs, driving under drugs, and over speeding.

6. RECOMENDATIONS

Authors recommended the following:

- The estimated models should be used to check the design of new or existing roads from the severity point of view.
- Concerned authorities should review the road maintenance schedule considering accident severity models into consideration.
- As the percentage of trucks increase in the traffic volume, authorities should plan to isolate it in a separate roads.
- Further research studies can utilize the economic point of view for comparing different means for fixing existing roads to minimize accident rates.
- A wider range of analysis types should be investigated to predict more severity models.
- Similar field tests should be applied on different roads types such as rural, agricultural and other major roads in Egypt.
- Other geometric road characteristics should be tested such as gradients, uphill or downhill travel, vertical curvature, super elevation, and traffic congestion.
- Examine the factors affecting speed variance, such as traffic composition and variable speed limits, so effective policies can be developed to reduce the speed variance. Similarly, driving behavior in congestion should be investigated so as to control and reduce aggressive driving in congested situations.

- Studies should be made to find out how to increase awareness among drivers in terms of accidents causes and how to be always in focus to drive safely.

REFERENCES

[1] WHO, world health statistics, world health organization, 2015.

[2] Central _Agency for Public Mobilization and Statistics, Egypt, 2015.

[3] El-Sayed A. "Accident Modeling and Valuation for Rural Roads in Egypt". M.Sc Thesis, Faculty of Engineering. Cairo University, Egypt, April 2002 .

[4] Abou El-Naga, I. M., "Accident Analysis on Main Roads Network of Dakahlia Governorate", M.Sc Thesis, Public Works Department, El-Mansoura University, 1998.

[5] Khaled A. Abbas, "Traffic safety assessment and development of predictive models for accidents on rural roads in Egypt" A paper accepted 2 December 2002 and published in Accident Analysis and Prevention (36), pp.l49-163, 2004.

[6] Behairy, W. A., "Study on Road Accidents Analysis and Prevention", M.Sc Thesis, Civil Engineering Department, Al-Azhar University, 1989.

[7] Babkov V. F., "Road Conditions and Traffic Safety", Published by Mir Publishers Moscow, 1989.

[8] Yousry, T. M., "Effect of Geometric Design and Traffic Characteristics on Highway Safety", M.Sc Thesis, Civil Engineering Department, Cairo University, 1992.

[9] The Relationship Between Traffic Congestion And Road Accidents: An Econometric Approach Using GIS Chao Wang February (2010).

[10] Zegeer C.V., R. Stewart, Council F., and Neuman T. "Accident Relationship of Roadway Width on Low-Volume Roads" A paper submitted for presentation at the 73th TRB, Washington, D.C, 1994.

[11] Jenkins D., "Car Driving before and after passing the driving test" Transport and Road Research Laboratory Report LR 499, 1979.

[12] Hrishikesh Kavade, A Logistic Regression Model to Predict Incident Severity Using The Human Factors Analysis and Classification System, M Sc thesis, the Graduate School of Clemson University, 2009

[13] Yuanchang Xie, Kaiguang Zhao, Nathan Huynh, Analysis of driver injury severity in rural single-vehicle crashes. Accident Analysis and Prevention 47 (2012) 36– 44, 2012.

[14] Yue Liu, Weather Impact on Road Accident Severity in Maryland. M Sc thesis, University of Maryland, College Park', 2013.

[15] Otte, D., Jänsch, M. and, Wiese, B, Injury Severity and Causation Factors of Motorcyclists in Traffic Accidents in comparing Drivers of Motorcycle and All Kinds of Motorized Two-wheelers. http://www.gidas.org, 2014.

[16] Train,. Discrete Choice Methods with Simulation. Cambridge University Press, 2002.

[17] Gwynn, D.W., Relationship of accident rates and accident involvements with hourly volumes, Traffic Quarterly. 21(3), pp. 407–418, 1967.

[18] Ceder, A. and Livneh, M., Relationships between road accidents and hourly traffic flow—I : analyses and interpretation. Accident Analysis & Prevention, 14(1), pp. 19-34, 1982.

[19] Hany M. Hassan, Nuha M. Albusaeedi, Atef M. Garib, Hussain A. Al-Harthei,. Exploring the nature and severity of heavy trucks crashes in Abu Dhabi. TRB 2015 Annual Meeting, 2015.

[20] Noland, R.B. and Oh, L.,. The effect of infrastructure and demographic change on traffic-related fatalities and crashes: a case study of Illinois county-level data. Accident Analysis & Prevention, 36(4), pp. 525-532, 2004.

[21] Jian Xu, Kara M. Kockelman, Yiyi Wang, MODELING CRASH AND FATALITY COUNTS ALONG MAINLANES AND FRONTAGE ROADS ACROSS TEXAS THE ROLES OF DESIGN, THE BUILT ENVIRONMENT, AND WEATHER. The 93rd Annual Meeting of the Transportation Research Board. 2014.

[22] Eluru, N., Bhat, C.R. and Hensher, D.A., A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis & Prevention, 40(3), pp. 1033-1054, 2008.

[23] Plainis, S., Murray, I.J. and Pallikaris, I.G., Road traffic casualties: understanding the night-time death toll. Injury Prevention, 12(2), pp. 125-138, 2006.