Journal of Traffic and Transportation Management
Volume 3, No. 1 (2021) pp. 11-16

Predicting Severity of Accidents in Malaysia By Ordinal Logistic Regression Models

Uneb Gazder[1*], Ashar Ahmed[2], Umaira Shahid[3]

[1] *Department of Civil Engineering, University of Bahrain, Isa Town, Bahrain, 32038*
[2] *Department of Urban and Infrastructure Engineering, NED University of Engineering and Technology, Karachi, Pakistan, 75270*
[3] *Independent Researcher*

Abstract

This study was aimed at determining the relationships of accident severity using road environment and traveller characteristics. Ordinal logistic regression models were used in this study. The accident data was provided by Malaysian Research Institute of Road Safety (MIROS) for all accidents which occurred in Penang state during 2006-2011. It was observed that motorbikes were predominantly involved in these accidents, hence, it was decided to develop three separate models; one for the overall data, and others for accidents with and without motorbikes. Logistic regression models showed that commercial land use, road width and experience of driver are important factors that may increase severity of accidents. Shoulder width was found to decrease the severity of motorbike accidents. Commercial land use, road width and driver experience have more impact on motorbike accidents as compared to accidents of other vehicles.

Keywords: *Traffic Accidents, Severity, Ordinal Logistic Regression, Motorbikes, Malaysia.*

1. Introduction

The need for transport and mobility has increased in the modern world. This has resulted in subsequent increase in motorization. One of the major issues related to motorization is traffic accidents [1]. These accidents are one of the leading causes of fatalities in the world, having severe economic impacts. According to World Health Organization (WHO), 65% of unintentional death-causing injuries are due to traffic accidents [2].
The contributing factors to a traffic accident can be categorized as follows: human, vehicle and environmental [3]. This study focuses on studying the impacts of environmental factors on traffic accidents involving different types of vehicles. Environmental factors may include, but not limited to, road surface, components of road geometry and light/visibility conditions [4]. Geometric factors which affect safety include, road cross-sectional reallocation, shoulder width, and lane width [5].
Most of the studies conducted on prediction of road traffic accidents have focused on human-related factors including: gender, age, marital status, driving experience and education [6]. The factors which have been reported to contribute to severity of accidents include: age, use of drugs and safety precautions (such as seatbelts and helmets) [7]. Other driver-related factors that have been observed to affect the severity of accidents are gender and marital status [8].
Similar to other low and middle-income countries, Malaysia also has growing concerns related to traffic accidents. It has been reported that the rate of deaths for road users is 24 per 100,000 people. A significant proportion of which occurs on the expressways [9].
It has been found in the previous literature that the accidents are predicted through binary logistic regression models or other prediction models. These models do not take in to account the concept of road characteristics that may increase the severity of road accidents. Therefore, this study employs ordinal logistic regression models, which have shared utility functions to account for variables increasing the severity of accidents. The use of these models has not been found in this area of research. Separate models have been developed for complete dataset as well as for accidents involving different types of vehicles. Previous studies have also been lacking in providing a comparative analysis of different vehicles along with geometric features of highways. Geometric features are important elements from the perspective of traffic engineering as they are easier to modify as compared to driver behavior. Furthermore, they inadvertently affect the drivers' behavior in turn. Hence, the results of this study are expected to be helpful, in reducing/mitigating traffic accidents, by providing design related solutions.

* Corresponding author. Tel.: +973 17876307
E-mail: unebgazdar@gmail.com

DOI: 10.5383/JTTM.03.01.002

2. Dataset and Analysis Method

Data for this study was collected by Malaysian Institute of Road Safety Research (MIROS) for the accidents occurring during the period of 2006-2011 in Penang state. Penang was selected because, as per the data acquired from MIROS, it had the highest vehicular density among all the states of Malaysia, as shown in Figure 1. Data related to 255 accidents was used from MIROS record. The measurements for road features were taken precisely through field observations. In addition to environmental features, victim-related characteristics were also recorded but this dataset was not found to be complete. So only those characteristics are included in this study which were recorded accurately and completely. However, the data was sufficient for this study as it only focuses on geometric and land use aspects' contribution to severity of accidents.

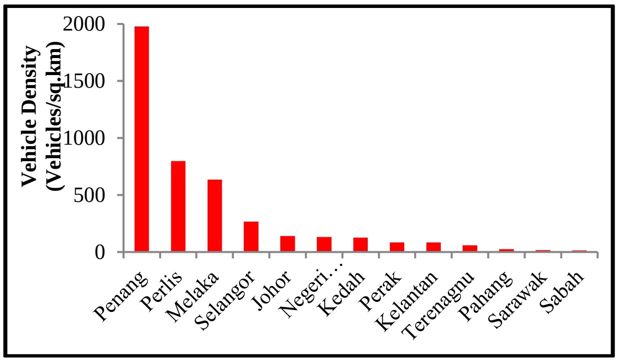

Fig. 1. Vehicle Density among the states of Malaysia

2.1. Description of Variables

The variables used in this study are shown in Table 1. These variables included road geometry and environment, location of accidents and few driver characteristics. There were several categorical variables, for which the number of categories were minimized by combining the categories with low frequencies. For e.g., most of the accidents occurred on four-leg or t-intersections, hence, other categories for type of highway were merged into a single category of 'others'. Therefore, if the variable for four-leg or t-intersection is '0' then it denotes 'others' category. The descriptive statistics of continuous variables are provided in Table 2.

Table 1. Variables collected for the accident analysis

Variable	Description
Accident severity	Categorical variable, 1 for minor injury accident, 2 for major injury accident and 3 for fatal accident
Four leg intersection	Categorical variable, 1 if the accident location was a four leg intersection
t-intersection	Categorical variable, 1 if the accident location was a t-intersection
Double lane marking	Categorical variable, 1 if the accident location had double line lane marking
Single lane marking	Categorical variable, 1 if the accident location had single line lane marking, lane marking variables will be 0 if there was no lane marking
Road width	Continuous variable, road width in meters
Shoulder width	Continuous variable, shoulder width in meters
Paved shoulder	Categorical variable, 1 if the accident location had paved shoulder
Speed limit	Continuous variable, speed limit in km/hr
Day	Categorical variable, 1 if the accident occurred in day timing
Night with light	Categorical variable, 1 if the accident occurred in night timing with street lights
Federal road	Categorical variable, 1 if the accident occurred on a federal road
State road	Categorical variable, 1 if the accident occurred on a state road
Municipal road	Categorical variable, 1 if the accident occurred on a municipal road
City	Categorical variable, 1 if the accident occurred in a city
Town	Categorical variable, 1 if the accident occurred in a town
Residential	Categorical variable, 1 if the accident occurred in a residential area
Commercial	Categorical variable, 1 if the accident occurred in a commercial area
Male	Categorical variable, 1 if the driver was male
Driver age	Continuous variable, drivers' age in years
License < 5	Categorical variable, 1 if driver had normal license for less than 5 years
License > 5	Categorical variable, 1 if driver had normal license for more than 5 years
NL	Categorical variable, 1 if driver had no license
L	Categorical variable, 1 if driver had learning license, all license variables will be 0 if driver had a police, military or international/foreign license
Primary education	Categorical variable, 1 is the driver had up to primary education
Secondary education	Categorical variable, 1 is the driver had more than primary education, both education variables will be '0' if driver does not have any education

Table 2. Descriptive statistics of continuous variables

Variable	Mean	Minimum	Maximum	Standard deviation
Road width (meters)	15.00	6	70	26
Shoulder width (meters)	1.18	0	20	2
Driver age	35.55	16	84	15

2.2. Ordinal Logistic Regression

Ordinal logistic regression can be used to predict variables which have ordered/ordinal responses. One of its applications include predicting severity of pain in the field of medical sciences [10]. This is similar to the problem for which these models are used in this study i.e. to predict severity of accidents. An alternative model to this problem could be multinomial logistic regression. However, fitting multinomial models requires large number of parameters and degrees of freedom. Moreover, it does not account for the ordering of response variable rendering its interpretation impractical [11].
Ordinal logistic regression model was proposed by McCullagh [12] who referred to it as proportional odds models. The probability of a response γ_i could be calculated as per Eq. 1.

$$\text{logit}(\gamma_i) = \theta - \beta(X) \tag{1}$$

Where, θ and β are the unknown model parameters and X is the input vector [13]. The right-hand side of the equation is also referred to as the utility function which is estimated during the development of the model using maximum likelihood approach. Ordinal models are considered simpler, compared to multinomial logistic, due to the fact that it works on the calculation of cumulative odds/probabilities as shown in Eq. 2.

$$logit(\gamma_i \le j) = \ln\left(\frac{p(\gamma_i \le j)}{1-p(\gamma_i \le j)}\right) \tag{2}$$

Where 'j' is the number of ordered categories for the response variable [14]. Figure 2 shows the application of logistic regression model.

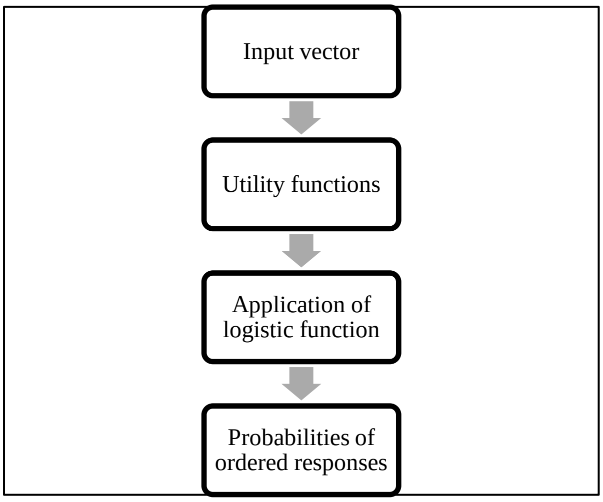

Fig. 2. Application logistic regression model

2.4. Involvement of Motorbikes

Motorbike riders have been considered as vulnerable road users as their involvement in accidents generally results in severe injuries or fatalities [23]. Keeping this fact in mind, it was decided to investigate if the involvement of motorbikes in road accidents for the Penang state is more than other vehicles. A statistical test was performed to check the null hypothesis that there is no significant difference in the percentages of accidents involving motorbikes and other vehicles. The critical t-value at 95% confidence level was 1.645. It was found that 76% of accidents involved other vehicles while 89% involved motorbikes. It should be noted that most of the accidents included motorbikes as well as other vehicles. T-value for the test was calculated as per Eq. 3 and Eq. 4 [24].

$$p = \frac{n_1 \times p_1 + n_2 \times p_2}{n_1 + n_2} \tag{3}$$

$$t = \frac{p_1 - p_2}{\sqrt{p(1-p)(1/(n_1+n_2)}} \tag{4}$$

Where, p_1 and p_2 represents proportion of accidents involving motorbikes and other vehicles, whereas n_1 and n_2 represent number of accidents involving motorbikes and other vehicles, respectively. It was found that the t-value for the test was found to be 7.10 which is higher than the critical t-value. Hence, the null hypothesis is rejected implying that traffic accidents involving motorbikes are significantly higher than those involving any other vehicle. It was further observed that 85% of all fatal accidents involved motorbikes. These findings confirm with other recent studies done in Malaysia which highlight the issue of high involvement of motorbikes in Malaysia [25, 26]. Moreover, number of motorbikes on Malaysian roads is increasing rapidly [26]. Considering these findings, it was decided to develop separate models for accidents with and without motorbikes to identify any specific factors affecting involvement of motorbikes.

3. Results and Discussion

Due to the significant involvement of motorbikes, there were three models developed in this study. First one was developed using complete dataset which is shown in Table 3. Second model was developed for accidents involving motorbikes, which is shown in Table 4 and the last one was developed using accidents involving other vehicles only (shown in Table 4). 25% of data was kept for testing all ordinal logistic models. Before developing the models, a correlation coefficient analysis was done for the all variables to check for multicollinearity and it was found that all coefficients were less than 0.8. Therefore, it was concluded that there is no multicollinearity between variables. Figure 3 provides a comparison of coefficients of those variables which were found to be common in all models.

Table 3. Ordinal logistic Regression model for Pooled data

	Estimate	**p-value**
Intercept 1	0.18	0.36
Intercept 2	-1.44	0.00
Single line lane marking	0.30	0.03
Commercial	0.57	0.00
License < 5	0.47	0.00
Road width	0.02	0.01
Initial Log likelihood	-280.146	
Final log likelihood	-253.298	
Accuracy (For training dataset)	57%	
Accuracy (For test dataset)	55%	

Table 4. Ordinal logistic Regression model for Motorbike Accidents

	Estimate	p-value
Intercept 1	0.08	0.48
Intercept 2	-1.49	0.00
Single line lane marking	0.44	0.01
Commercial	0.40	0.00
License < 5	0.37	
Road width	0.08	0.00
Shoulder width	-0.52	0.00
Initial Log likelihood	-245.035	
Final Log likelihood	-214.604	
Accuracy (For training dataset)	61%	
Accuracy (For test dataset)	58%	

Table 5. Ordinal logistic Regression model for Other Vehicle Accidents

	Estimate	p-value
Intercept 1	0.55	0.00
Intercept 2	-1.25	0.00
Commercial	0.47	
License < 5	0.54	0.02
Road width	0.02	0.00
Initial Log likelihood	-211.483	
Final Log likelihood	-195.804	
Accuracy (For training dataset)	61%	
Accuracy (For test dataset)	60%	

The model in Table 3 shows that accident severity is positively affected by lane marking, commercial land use and road width among the roadway environment factors. This is in agreement with the literature [26], which states that wider roads appear to be less dangerous; thus, promoting risky behavior. The only factor related to driver which significantly affected the severity was having driving license for less than 5 years which affected the severity positively. Single line lane markings were present on roads having less traffic volume, as compared to double line lane markings, and more road width encouraged the driver to drive at a higher speed, thus both these variables contributed positively towards severe accidents. The traffic demand in commercial areas may include heavy vehicles which add to the severity of accidents. Drivers having license for less years are less experienced and may be involved in dangerous maneuvers and thus having severe accidents.

The model shown in Table 4 shows almost the same results as the pooled data. However, shoulder width has also been found to have negative impact on motorbike accident severity. The reason for this could be that motorbike drivers tend to use the wide shoulders more often for making their passing (overtaking) maneuvers creating merging and diverging patterns. More shoulder width allows them to do these maneuvers with less interference to the travel lane, which may result in less severe accidents. Moreover, shoulders also allow these maneuvers to be less severe than head-on or right-angle accidents.

Table 5 shows that the severity of accidents for other vehicles is not affected by lane marking which was a significant factor for the pooled data and motorbike accidents. Hence, it can be said that the drivers of other vehicles tend to follow their lane/path/direction more regularly than motorbike drivers. Therefore, presence of lane marking is not an important factor in their accidents. Other variables have the same relationship with these accidents as that for pooled data (as shown in Table 3).

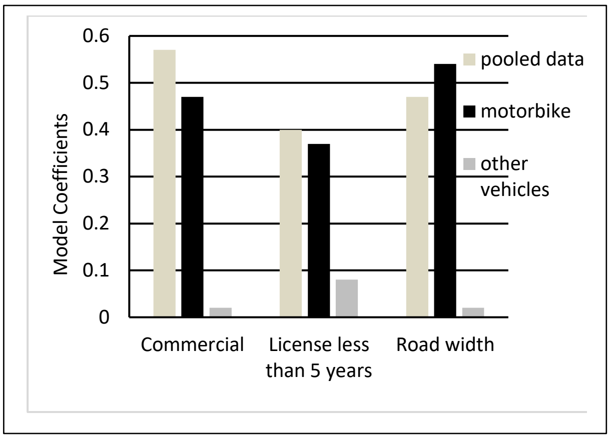

Fig. 3. Application of logistic regression model

4. Conclusion

This study focused on finding the impacts of road geometry and environmental factors on severity of accidents in Penang state of Malaysia. Moreover, effects of these variables on accidents with and without motorbikes were also compared. Ordinal logistic regression models were used for this purpose and their application in this field is not found in the previous literature. It was found that motorbikes involvement in accidents is significantly higher than other vehicles having more than 80% of fatal accidents.

Separate models were developed for pooled data, accidents with motorbikes and accidents with other vehicles. It was shown that commercial land use, road width and experience of driver are important factors that add to severity of accidents in all cases. Shoulder width was found to have a negative impact on severity of motorbike accidents, possibly because it provides more space for motorbike riders to do merging or diverging maneuvers. Lane marking was not found as important parameter for accidents with other types of vehicles. Possible reason for this could be the fact that car or heavy vehicle drivers tend to follow the path more appropriately than motorbike drivers in spite of single lane marking. Commercial land use, road width and driver experience have more impact on motorbike accidents as compared to other vehicles.

In light of the results from this study, it is recommended that licensing regulations should be made more stringent to avoid severe accidents by inexperienced drivers. Moreover, relevant authorities should make more stringent regulations to encourage drivers (especially for motorbikes) to follow their path. The road agencies should take actions that compel drivers to follow regulations, such as providing double lane markings. Highways in commercial areas should be provided with more shoulder width to avoid severe motorbike accidents.

Acknowledgments

The authors are obliged to Prof. Dr. Ahmad Farhan Mohd Sadullah, Professor of Transportation Engineering at the School of Civil Engineering, Universiti Sains Malaysia, for his contribution in data collection. The authors duly acknowledge the Malaysian Institute of Road Safety Research (MIROS) for providing accident data for this research.

References

[1] Zhang, G., K. K. Yau, X. Zhang, and Li, Y. Traffic accidents involving fatigue driving and their extent of casualties. Accident Analysis and Prevention 2016; 87: 34-42. https://doi.org/10.1016/j.aap.2015.10.033

[2] Hasanzadeh, J., M. Moradinazar, F., Najafi, and Ahmadi-Jouybary, T. Trends of mortality of road traffic accidents in Fars Province, Southern Iran, 2004-2010.Iranian Journal of Public Health 2014; 43(9): 1259.

[3] Shankar, V., and Venkatesan, M. Pattern of injuries due to fatal road traffic accidents from rural Chennai. Online International Journal of Medical and Social Sciences 2015; 1(2): 30-36.

[4] Lankarani, K. B., S. T. Heydari, M. R. Aghabeigi, G. Moafian, A. Hoseinzadeh, and Vossoughi, M. The impact of environmental factors on traffic accidents in Iran. Journal of Injury &Violence Research 2014; 6(2): 64–71. https://doi.org/10.5249/jivr.v6i2.318

[5] Bassat, T.B., and Shinar, D. Effect of shoulder width, guardrail and roadway geometry on driver perception and behavior. Accident Analysis and Prevention 2011; 43(6): 2142-2152. https://doi.org/10.1016/j.aap.2011.06.004

[6] Raina, P., N. Sohel, M. Oremus, H. Shannon, P. Mony, R. Kumar, and Yusuf, R. Assessing global risk factors for non-fatal injuries from road traffic accidents and falls in adults aged 35–70 years in 17 countries: a cross-sectional analysis of the Prospective Urban Rural Epidemiological (PURE) study.Injury Prevention 2016; 22(2): 92-98. https://doi.org/10.1136/injuryprev-2014-041476

[7] Mohamad, N., M. Muhammad, M. Haque, Z. K. Jawad, N. H. Abu Bakar, S. Ismail, and Simbak, N. Traumatic motor vehicle accidents of malaysia: implications of illicit drugs use. International Medical Journal, Japan 2016; 23(2): 192-194.

[8] Nordin, R., N. Abdul Rahman, M. F. Rashdi, A. Yusoff, R. Abdul Rahman, S. Sulong, A. M. Nur, N. Abdul Razak, M. N. Abdul Jabar, M. I. Ibrahim, N. Mohamed, S. H. R. Ahmad, and Ramli, R. Oral and maxillofacial trauma caused by road traffic accident in two university hospitals in Malaysia: A cross-sectional study. Journal of Oral and Maxillofacial Surgery, Medicine, and Pathology 2015; 27(2): 166-171. https://doi.org/10.1016/j.ajoms.2014.01.001

[9] Sameen, M. I. and Pradhan, B. Severity prediction of traffic accidents with recurrent neural networks. Applied Sciences 2017; 7(6): 476. https://doi.org/10.3390/app7060476

[10] Harrell, F. E. Ordinal logistic regression. Regression modeling strategies, Springer 2001; New York, NY: 331-343. https://doi.org/10.1007/978-1-4757-3462-1_13

[11] Warner, P. Ordinal logistic regression. BMJ Sexual and Reproductive Health 2008; 34(3): 169-170. https://doi.org/10.1783/147118908784734945

[12] McCullagh, P. Regression models for ordinal data. Journal of the royal Statistical Society 1980; Series B (Methodological): 109-142. https://doi.org/10.1111/j.2517-6161.1980.tb01109.x

[13] Brant, R. Assessing proportionality in the proportional odds model for ordinal logistic regression. Biometrics 1990; 1171-1178. https://doi.org/10.2307/2532457

[14] Bender, R., and Grouven, U. Ordinal logistic regression in medical research. Journal of the Royal College of physicians of London 1997; 31(5): 546-551.

[15] Maier, H. R. and Dandy, G. C. The Use of Artificial Neural Networks for The Prediction of Water Quality Parameters. Water Resources Research 1996; 32(4): 1013-1022. https://doi.org/10.1029/96WR03529

[16] Zhang, G., Patuwo, B. E. and Hu, M. Y. Forecasting with Artificial Neural Networks: The State of The Art. International Journal of Forecasting 1998; 14(1): 35-62. https://doi.org/10.1016/S0169-2070(97)00044-7

[17] Imrie, C. E., Durucan, S., and Korre, A. River Flow Prediction Using Artificial Neural Networks: Generalisation Beyond the Calibration Range. Journal of Hydrology 2000; 233(1): 138-153. https://doi.org/10.1016/S0022-1694(00)00228-6

[18] Duan, Z. H., Kou, S. C., and Poon,C. S. Prediction of Compressive Strength of Recycled Aggregate Concrete Using Artificial Neural Networks. Construction and Building Materials 2013; 40: 1200-1206. https://doi.org/10.1016/j.conbuildmat.2012.04.063

[19] Thompson, M. L. and Kramer, M. A. Modeling Chemical Processes Using Prior Knowledge and Neural Networks. AIChE Journal 1994; 40(8): 1328-1340. https://doi.org/10.1002/aic.690400806

[20] Zealand, C. M., Burn, D. H., and Simonovic, S. P. Short Term Streamflow Forecasting Using Artificial Neural Networks. Journal of Hydrology 1999; 214(1): 32-48. https://doi.org/10.1016/S0022-1694(98)00242-X

[21] Hsu, K. L., Gupta, H. V., and Sorooshian, S. Artificial Neural Network Modeling of The Rainfall-Runoff Process. Water Resources Research 1995; 31(10): 2517-2530. https://doi.org/10.1029/95WR01955

[22] Srivastava, N., Hinton, G. E., Krizhevsky,A., Sutskever, I., and Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research 2014; 15(1): 1929-1958.

[23] Di Stasi, L. L., D. Contreras, A. Cándido, J. J. Cañas, and Catena, A. Behavioral and eye-movement measures to track improvements in driving skills of vulnerable road users: First-time motorcycle riders. Transportation research part F: traffic psychology and behaviour 2011; 14(1): 26-35. https://doi.org/10.1016/j.trf.2010.09.003

[24] Zanella, A., R. C. Silveira, L. F. Roesch, A. L. Corso, P. T. Dobbler, V. Mai, and Procianoy, R. S. Influence of own mother's milk and different proportions of formula on intestinal microbiota of very preterm newborns. PloS one 2019; 14(5): e0217296. https://doi.org/10.1371/journal.pone.0217296

[25] Omar, W. N. S. B. Speed studies on Malaysia East Coast expressway. Thesis submitted in fulfillment of the requirements for the award of the degree of Bachelor Engineering (Hons.) Civil Engineering, Faculty of Civil Engineering & Earth Resources, Universiti Malaysia Pahang 2017.

[26] Abdullah, M. N., M. A. M. Shukran, M. N. Ismail, K. Maskat, M. R. M. Isa, and Ishak, M. S. The study of technological prevention method of road accident related to driver and vehicle. Journal of Fundamental and Applied Sciences 2017; 9(35): 621-629. https://doi.org/10.4314/jfas.v9i3s.48