Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

A LOGISTIC REGRESSION ANALYSIS ON THE INFLUENCE OF ACCIDENT FACTORS ON THE FATALITIES OF ROAD ACCIDENTS IN METRO MANILA

J. D. Urrutia[1,*], S. D. Bobihis[2], C. D. Serrano[2], J. Mercado[3]and R. R. Bernardino[4]

[1]Director, Intellectual Property and Management Office, Office of the Vice President for Research, Extension, and Development

[2]Polytechnic University of the Philippines

[3]Vice President for Research, Extension, and Development, Polytechnic University of the Philippines

[4]Faculty Member, College of Science, Polytechnic University of the Philippines – Sta. Mesa, Manila

Published online: 24 February 2018

ABSTRACT

Using binary logistic regression, the data collected from Metropolitan Manila Development Authority (MMDA) were examined to determine if accident factors contribute to accident fatality. A total of 29,409 accidents were studied after tallying all the accidents with the complete information on all the considered factors. The classification (fatal or nonfatal) is the dependent variable while the other factors such as age, gender, junction-type, weather, location, time and vehicle-type are the independent variables. The model formulated in this study shows the relationship of the dependent variable and the significant factors in terms of odds ratio concept. The findings show that all the independent variables except for age are statistically significant in the accident fatality of road accidents in Metro Manila from 2006 to 2016.

Keywords: Binary logistic regression, Odds ratio, Probabilities, Regression coefficients, Regression model

Author Correspondence, e-mail: jackieurrutia20@gmail.com

doi: http://dx.doi.org/10.4314/jfas.v10i3s.4

INTRODUCTION

Accident is a phenomenon which humans cannot escape. One type of accident is the road accident, which occurs when a vehicle collides with another vehicle, pedestrian, animal, road debris, or other stationary obstruction, such as a tree or utility pole. Road accident is classified as the fourth leading cause of death in the Philippines according to the World Health Organization (WHO) (Peden, 2004).

Identifying the most probable factors that affect accident severity can help in preventing accidents and reducing its fatality. In a study in Riyadh, they estimated the effect of the statistically significant factors on accident severity and using a data set that is derived from a sample of 560 subjects involved in serious accidents re-ported in traffic police records (Al-Ghamdi, 2002).

In another study, they assessed the risk factors on accident fatality. The variable accident severity is considered as a dichotomous response variable, and the explanatory variables such astime, location, type of vehicle, gender, license status, cause of accident, and type of accident are treated as influencing factors on the accident severity. A series of statistical analyses were conducted, and the results revealed that the model can be used for the safety improvements against the traffic accidents in Jaffna, Sri Lanka (Renuraj, 2015).

Road accidents claim lives and cause disabilities put significant strain on families. For every person killed, injured or disabled, many others are deeply affected. Affected families are driven into poverty as a result of the high cost of prolonged medical care, the loss of a family breadwinner, or the extra funds needed to care for people with disabilities (Ala, 2008).

Metropolitan Manila Accidents Reporting and Analysis System (MMARAS) reported that 861,744 accidents had happened from January 2005 to November 2016. These resulted in 4,126 fatality accidents which involve at least one person killed, 160,095 non-fatal accidents where at least one person was injured but no fatalities, and 697,552 incidents of damaged to properties. These reveal that an average of 78,340 accidents happens every year and 375 people lose their lives due to these accidents.

The main objective of this study is to examine some factors which believed to have a higher potential for serious injury and death such as the gender of the person involved, location of the accident, the time of day, as well as the weather, when the accident occurred, the type of vehicle involved in the accident, the junction-type where the accident happened and the causes of those accidents. Using binary logistic regression, the researchers formulate the best model to estimate the influence of these factors on the fatalities of road accidents in NCR.

RESEARCH PARADIGM

A research paradigm was constructed in conducting this study indicating the inputs, process and outputs that the researchers used.

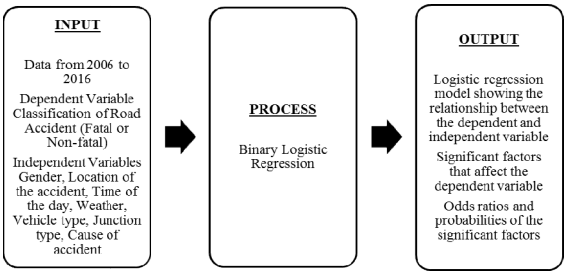

Fig.1. Research paradigm

The independent variables are the gender, location of the accident, time of the day the accident occurred, vehicle-type involved in the accident, junction-type and the cause of the accident.

Each variable is divided into different categories. For gender, it was either male or female. For the location of the accident, the district is divided into five such as Northern District, Western District, Eastern District, Southern District and Central District, which are the districts of the Metropolitan Manila.

The time of day the accident occurred is based from Land Transportation and Franchising Regulatory Board's (LTFRB) definition of rush hour and these are 10:01pm-4:00am are the times before morning rush hour, 4:01am-10:00am is considered as morning rush hour, 10:01am-4:00pm are the times before evening rush hour, and 4:01pm-10:00pm is considered as evening rush hour. Junction types were classified as Crossroad, Rotunda, Junctions (K-Junction T-Junction, Y-Junction and Skew Junction), U-turn slots and Not Junction.

Vehicle Types were categorized based on the classification of MMARAS and these are Bicycle/Pedicab (human-powered vehicle), Motorcycle (two-wheeled mechanically propelled Vehicle), Motor Tricycle (three-wheeled passenger-carrying mechanically propelled vehicle), Car (privately-owned mechanically propelled vehicle, which included all forms of 'Private use' small passenger-carrying vehicles), Jeepneys/Taxi/Fx/Bus (mechanically-propelled vehicle which carries passengers on payment of a fee), Van (small vehicle for carrying goods), Truck (large vehicle for carrying goods) and Train. The dependent variable is the classification, that is Fatal or Non-Fatal.

STATEMENT OF THE PROBLEM

Aside from answering the main objective of this study, the researchers also aim to answer the following questions:

1. What are the numbers of accidents in Metro Manila from 2006 to 2016?

a. Gender

b. Location

c. Time of day

d. weather

e. Vehicle type

f. Junction type

g. Classification

h. Causes

2. What is the logistic regression model for accident fatality of road accidents in Metro Manila?

3. What are the factors which is significant in the accident severity of road accidents in NCR?

4. What is the odds ratio and probability of each significant factors and its effect to the outcome of accident fatality?

SCOPE AND LIMITATION

The study was conducted in National Capital Region (NCR), a metropolitan region composed of the City of Manila and the surrounding cities of Caloocan, Las Piñas, Makati, Malabon, Mandaluyong, Marikina, Muntinlupa, Navotas, Parañaque, Pasay, Pasig, Quezon City, San Juan, Taguig, and Valenzuela, as well as the Municipality of Pateros.

The National Capital Region has been focused to be the area of the study since it is considered as political, economic, social, cultural and educational center of the Philippines and because of the accessibility and availability data gathered by Metropolitan Manila Development Authority (MMDA). The list of data gathered was composed of the number of accidents, date, place, and time it happened, as well as the number of fatalities, causes of accidents, vehicle types involved and age and gender of the person involved in the accident from January 2006 to December 2016.

MATERIALS AND METHODS

The researchers used a descriptive type of research for this study. The study reveals the status of road accidents in National Capital Region. A specialized software, Matlab, was used to analyze the data gathered.

Binary Logistic Regression

Binary logistic regression estimates the probability that a characteristic is present (e.g. estimate probability of success) given the values of explanatory variables. The model is formulated by means of:

$$\pi_i = \frac{\exp(\beta_0 + \beta_1 x_1}{1 + \exp(\beta_0 + \beta_1 x_i)} \qquad (1)$$

$$\log it(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) \qquad (2)$$

The road accident is the binary response variable – 1 if the trait is present in observation; 0 if it is not. The independent variables are the explanatory variables which can be discrete, continuous, or a combination.

There are several assumptions before performing a logistic regression:

1. The data Y_1, Y_2, ... Y_n are independently distributed, i.e., cases are independent.
2. Distribution of Y_i is $B_{in}(n_i, \pi_i)$, i.e., binary logistic regression model assumes binomial distribution of the response. The dependent variable does NOT need to be normally distributed, but it typically assumes a distribution from an exponential family (e.g. binomial, Poisson, multinomial, normal,)
3. Does NOT assume a linear relationship between the dependent variable and the independent variables, but it does assume linear relationship between the logit of the response and the explanatory variables; $\text{logit}(\pi) = \beta_0 + \beta_x$.
4. Independent (explanatory) variables can be even the power terms or some other nonlinear transformations of the original independent variables.
5. The homogeneity of variance does NOT need to be satisfied. In fact, it is not even possible in many cases given the model structure.
6. Errors need to be independent but NOT normally distributed.

7. It uses maximum likelihood estimation (MLE) rather than ordinary least squares (OLS) to estimate the parameters, and thus relies on large-sample approximations.

8. Goodness-of-fit measures rely on sufficiently large samples, where a heuristic rule is that not more than 20% of the expected cells counts are less than 5.

RESULTS AND DISCUSSIONS

This chapter focuses on the preliminary analysis of data on the number of road accidents in Metro Manila and the detailed analysis of the data using the statistical tool, Binary Logistic regression, in software called MatLab to fit a model to the data.

Behavior of Road Accidents in Metro Manila

The data obtained from the MMDA was thoroughly tallied because some important values for the independent variables were missing. So, to properly apply the logistic regression, we only get those parts with complete and understandable data to avoid confusion. After this, we have come up with a total of 29,409 accidents for the year 2006 to 2016.

Table 1. The Frequency and Percentage of Road Accidents in Manila Based on the Gender of the Person involved in the Accident

Gender	Frequency	Percentage
Male	21706	73.80
Female	7703	26.20
Total	29409	100

Table 2. The Frequency and Percentage of Road Accidents based on its Location

District	Frequency	Percentage
Eastern	4617	15.70%
Western	1108	3.80%
Northern	4285	14.60%
Southern	9482	32.20%
Central	9917	33.70%
Total	29409	100 %

Table 3. The Frequency and Percentage of Road Accidents based on the Time of Day

Time of the Day	Frequency	Percentage
Morning Rush Hour	7937	27 %
Before Evening Rush Hour	9183	31.20%
Evening Rush Hour	8363	28.40%
Before Morning Rush Hour	3926	13.30%
Total	29409	100 %

Table 4. The Frequency and Percentage of Road Accidents based on the Weather

Weather	Frequency	Percentage
Dry	10717	36.40%
Wet	354	1.20%
Fair	18338	62.40%
Total	29409	100 %

Table 5. The Frequency and Percentage of Road Accidents in terms of the Type of Vehicle

Vehicle Type	Frequency	Percentage
Bike/Pedicab	1348	4.60%
Motorcycle	10897	37 %
Motor Tricycle	1516	5.20%
Car	7994	27.20%
Jeepney	2939	10 %
Fx/Taxi	1764	6 %
Bus	825	2.80%
Van	1090	3.70%
Truck	1023	3.50%
Train	13	0%
Total	29409	100 %

Table 6. The Frequency and Percentage of Road Accidents in terms of the Junction Type from

Junction Type	Frequency	Percentage
Crossroads	2244	7.60%
Junctions	1852	6.30%
Rotunda	419	1.40%
U-turn Slots	500	1.70%
Not a junction	24394	82.90%
Total	29409	100 %

Table 7. The frequency and Percentage of Road Accidents in terms of its Classification

Classification	Frequency	Percentage
Fatal	3903	13.30%
Nonfatal	25506	86.70%
Total	29409	100 %

Table 8. The Frequency and Percentage of Road Accidents in terms of its Causes

Causes	Frequency	Percentage
Human Error	7546	25.70%
Environmental Error	1483	5 %
Mechanical Error	1	0 %
No Accident Factor	20379	69.30%
Total	29409	100 %

Logistic Regression Model

A binary logistic regression model was formulated by the researchers to estimate the influence of the explanatory variables to the road accidents happening in NCR. Using the 18 considered factors of road accidents, the regression coefficients were obtained and the model was written in the form of:

$$y = e^{\begin{array}{l}1.402+0.337x_1-0.423x_2-0.626x_3+0.333x_4-2.983x_5+0.609x_6\\+0.258x_7+0.563x_8+0.459x_9-0.875x_{10}-0.583x_{11}+0.364x_{12}\\+0.361x_{13}+2.796x_{14}-1.050x_{15}-0.740x_{16}+3.194x_{17}-3.884x_{18}\end{array}}$$

where y is the level of fatality,x_1 is jeepney, x_2 is van, x_3 is U-turn slots, x_4 is Northern district, x_5 is truck, x_6 is FX or taxi, x_7 is female, x_8 is car, x_9 is before morning rush, x_{10} is train, x_{11} is not a junction, x_{12} is evening rush hour, x_{13} is before evening rush hour, x_{14} is wet, x_{15} is junctions, x_{16} is no accident factor,x_{17} is fair and x_{18} is Mechanical error.

Factors of Severity of Road Accidents in NCR

The following tables show the individual test of significance of each Regression Coefficients at 0.01 level of significance.

Table 9. Individual Test of Significance of Regression Coefficients (District)

Factors	β	p-value
Western District	0.135	0.218
Northern District	0.333	0.000
Southern District	0.126	0.047
Central District	-0.121	0.049

Table 10. Individual Test of Significance of Regression Coefficients (Time)

Factors	β	p-value
Before evening rush hour	0.361	0
Evening rush hour	0.364	0
Before morning rush hour	0.459	0

Table 11. Individual Test of Significance of Regression Coefficients (Junction)

Factors	β	p-value
Junctions	-1.050	0
Rotunda	-0.164	0.547
U-turn slots	-0.626	0.001
Not a junction	-0.583	0

Table 12. Individual Test of Significance of Regression Coefficients (Weather)

Factors	β	p-value
Wet	2.796	0
Fair	3.194	0

Table 13. Individual Test of Significance of Regression Coefficients (Causes)

Factors	β	p-value
Mechanical error	-3.884	0
Environmental error	16.787	1.000
No accident factor	-0.740	0

Table 14. Individual Test of Significance of Regression Coefficients (Vehicle)

Factors	β	p-value
Motorcycle	0.216	0.039
Tricycle	0.321	0.016
Car	0.563	0
Jeepney	0.337	0.004
FX/taxi	0.690	0
Bus	-0.252	0.113
Van	-0.423	0.003
Train	-0.875	0
Truck	-2.983	0
Female	0.258	0
Constant	1.402	0

It is revealed that western district, southern district, central district, rotunda, mechanical error, motorcycle, tricycle and bus are not statistically significant since their computed p-value are greater than the 0.01 level of significance. The rest of the factors in tables are all significant since their p-values are all less than 0.01.

Odds Ratio and Probabilities

By omitting the insignificant factors on Table 9, we have observed that eighteen of the considered factors have significance in the accident fatality of road accidents as shown in Table 15.

Table 15. Odds ratio and Probability of the Factors

Factors	β	p-value	Odds ratio	Probabilities
Jeepney	0.337	0.004	1.401	0.583
Van	-0.423	0.003	0.655	0.396
U-turn slots	-0.626	0.001	0.535	0.348
Northern District	0.333	0.000	1.395	0.583
Truck	-2.983	0.000	0.051	0.048
Fx/taxi	0.690	0.000	1.993	0.666
Female	0.258	0.000	1.294	0.564
Car	0.563	0.000	1.757	0.637
Before morning rush hour	0.459	0.000	1.582	0.613
Train	-0.875	0.000	0.417	0.294
Not a junction	-0.583	0.000	0.558	0.358
Evening rush hour	0.364	0.000	1.439	0.590
Before evening rush hour	0.361	0.000	1.435	0.589
Wet	2.796	0.000	16.384	0.942
Junctions	-1.050	0.000	0.350	0.259
No accident factor	-0.740	0.000	0.477	0.323
Fair	3.194	0.000	24.382	0.961
Mechanical error	-3.884	0.000	0.021	0.020

The odds of having a nonfatal kind of accident while driving or riding a jeepney is 1.401 higher than when using a bike/pedicab(baseline for vehicle) and a probability of 0.583 to have a positive outcome. The odds of having a nonfatal outcome if the person involved in the accident is female is 1.294 higher than in male with probability equal to 0.564. The baseline categories for district, time, junction, weather and causes are eastern district, morning rush hour, crossroads, dry and human error, respectively.

CONCLUSIONS

Since the dependent variable in this study is binary (fatal or nonfatal), binary logistic regression is used in the formulation of the model which estimates the influence of accident factors in the level of fatality of road accidents in Metro Manila from 2006-2016. Seven independent variables were used in the model development process prior to the data gathered from the Metropolitan Manila Development Authority (MMDA).

By setting a 0.01 level of significance and performing some tests in SPSS and MatLab, all of the independent variables were included in the model such as the gender, district (location), time, junction type, weather, causes of accidents and vehicle types. The odds presented in the results of this paper can be used in establishing priorities for programs to reduce serious accidents in the near future.

REFERENCES

Ala, V. (2008). Road Safety as a Public Health Priority. Department of Health. See also URL http://www.doh.gov.ph/sites/default/files/publications/Vol3Issue3November2008.pdf

Al-Ghamdi, A. (2002). Using Logistic Regression to Estimate the Influence of Accident Factors on Accident Severity. See also URLhttps://www.ncbi.nlm.nih.gov/pubmed/12371778

Dovom, H. (2012). An Analysis of Pedestrian Fatal Accident Severity Using a Binary Logistic Regression Model. ITE Journal. See also URL trid.trb.org/view.aspx?id=1218386

Haadi, A. (2014). Identification of Factors that Cause Severity of Road Accidents in Ghana: A Case Study of the Northern Region. International Journal of Applied Science and Technology. See also URL www.ijastnet.com/journals/Vol_4_No_3_May_2014/28.pdf

Peden, M. (2004). World Report on Road Traffic Injury Prevention. World Health Organization. See also URL apps.who.int/iris/bitstream/10665/42871/1/9241562609.pdf

Renuraj S. (2015). Factors Influencing Traffic Accidents in Jaffna. See also URL http://doi.org/10.4038
/sljastats.v16i2.7827

How to cite this article:
Urrutia J D, Bobihis S D, Serrano C D, Mercado J, Bernardino R R. A logistic regression analysis on the influence of accident factors on the fatalities of road accidents in metro manila. J. Fundam. Appl. Sci., 2018, *10(3S), 32-44.*