Modeling of Road Accidents in Penang Island, Malaysia

Allam Musbah Al Allam[1,*], Essa Altboli[2], Fthei Ramadan Lamein[3], Hosni Abdulruhman Saleh[4]

[1]Department of Civil and Structural Engineering, Universiti Kebangsaan Malaysia, Selangor, Malaysia
[1]Engineering & Information Technology Research Center, Bani walid, Libya
[1,2,3,4]Department of Civil Engineering, College of Technical Sciences. Bani walid, Libya
[3]Department of Civil Engineering, Universities Muhammadiyah Surakarta, Indonesia

alallm84@yahoo.com*

***Abstract*-- Road accidents are in an increasing rate worldwide and in Malaysia. In Malaysia, news cars are registered and being used in the streets thus, this will result in congestion and high rates of accidents. Based on Malaysian statistics, every 11 minutes one accident reported and at least one person dies every 2 and half hours in Malaysia. This paper aims to develop a model which represents the relationship between the time and the number of dead people in the Penang Island in Malaysia. The model developed shows very high correlation between the time and the number of dead people which is equal to 0.779.**

***Keywords*--Modelling; Penang Island; road accidents.**

I. INTRODUCTION

STATISTICS from the world health organization indicates that approximately 3000 people die from road traffic accidents around the world each day and two million people die worldwide every year [1]. Therefore, a road traffic accident involves high human suffering and socioeconomic costs in terms of premature deaths, injuries, loss of productivity. During 2008, road traffic injuries ranked fourth among the leading causes of death in the world. Globally, we are now striving to eradicate communicable diseases, control the no communicable diseases, and injuries[2]. Moreover, Hayley & Stough [3] found that driving, under the influence of alcohol is known to result in a significantly increased risk for being severely injured or killed as a result of road traffic accident. Despite the frequency of traumatic injuries due to road accidents and potential importance of identifying children at risk of impaired recovery one year after a road accident [4]. Therefore, road traffic networks are key economic drivers in today's world. They provide a quick, reliable and flexible transportation system, for people, goods and services [5]. According to, Chung & Chang [6], found that, the process of roadway safety improvement is based on the collection and analysis of traffic accident data. In general, highly sophisticated statistical and mathematical models are applied to identify accident causal factors by using the collected accident data and its associated data such as traffic data, demographical data, and others. As pointed by [7], the road accidents can be caused by different factors such as human factors. Quality of the decision-making process of drivers could have a considerable impact on preventing disasters. Identification of the most relevant factors for explaining road accident occurrence is an important issue in road safety research, particularly for future decision-making processes in transport policy. As reported by Dadashova, B., et.l. [8], the road accidents are very complex processes which can be explained as a consequence of set of different influential factors. Impact of the external factors on road accidents is studied through statistical and time series models. Meanwhile, [9] demonstrated that, the modeling road accident occurrence has gained increasing attention over the years. So far, considerable efforts have been made from researchers and policy makers in order to explain road accidents and improve road safety performance of highways. Therefore, accidents impose serious problems to society in terms of human costs, economic costs, property damage costs and medical costs. Understanding the various factors that affect accident occurrence is of particular concern to decision makers and researchers. Xu et al. [10] are mentioned that, developed accident risk models for different traffic states. Traffic flow parameters were found to have different effects on safety for every traffic state. For instance, the average downstream occupancy seemed to reduce accident risk in two traffic states (in congested traffic as well as in transition from free flow to congested flow) but caused an increase in the overall model. Nowadays, road accidents are a major public health problem, which increase is forecasted if road safety is not treated properly, dying about 1.2 million people every year around the globe [11]. Currently, the raw data collection is not a problem, but the ability to identify methods and models that can transform data into reliable information or knowledge is a challenge. As the capacity to collect and store data increases faster than our ability to make decisions based on that data, most of the time the data are stored without properly filtering, refinement and processing [12]. On the other hand, the road traffic accidents are alarming but ignored public health challenge that requires concrete efforts for effective and viable prevention [13]. Meanwhile, [14], are found that, the road traffic accident (RTA) is the most common cause of injuries and currently the fifth principal cause of death in Malaysia with 5.8%. Road accidents are essentially caused by improper interactions between vehicles, between vehicles and other road users and/or roadway features. Road accident is a serious problem in Malaysia. In year 2008 there were 373,047 road accidents that lead to 6,527 fatalities. The number of fatalities due to road

accidents in Malaysia has consistently been above 6,000 since year 2003, whilst the indices used for road safety indicator has shown only slight decline over the years. Due to this alarming figure, the government has launched the Road Safety Plan 2006-2010 in March 2006 with the objectives to reduce fatality rates to 2 fatalities per 10,000 registered vehicles, 10 fatalities per 100,000 populations and 10 fatalities per billion vehicles kilometer travelled (VKT) by year 2010 Ministry of Transport, (2006). Collection of quality, accurate and reliable data that are collected over a period of time are needed in order to understand the factors influencing the arising figures of road accidents and injuries [15][16]. In addition, advanced road accident analysis system is needed to help strategies road safety initiative as well as inculcate better understanding of road accident causation. Furthermore, accident data is critical to monitor and evaluate the effectiveness of road safety interventions introduced by the government and road authorities. In 1995, the Road Safety Research Centre, at Universiti Putra Malaysia was mandated with the task of reviewing the earlier national safety target. The outcome of the research [17] was the statistical model whichpredicted some 9,127 deaths in the year 2000 if traffic exposure continued to increase at the projected rate. Accidents statistical by Jorgensen 1978 [18] in 2001 to 2003 shows that estimated 46% to 49% fatal and serious accident occurred at State and Municipal roads. Current practice of road safety improvement only on Federal road and with an earlier finding, there's new type of road need to be tackle as to improve road safety for all road users. In order to accomplish the target, Cabinet Committee on Road Safety on 25 May 2004 have decided that road safety audit and road safety programmers to be implemented onto State Road and Municipal Road. In 9th Malaysian Plan, allocation of about MYR 200 million was provided by the Government to improve the highest 5 hazardous locations along state and municipal roads [19]. With the recent endorsement of the National Road Safety Blueprint 2001-2010, strategic activities have prioritized and key performance indicators have been established. They are monitored regularly by the Road Safety Department Malaysia and will be regularly reported to the Cabinet Committee on Road Safety chaired by the Prime Minister of Malaysia [20].

II. METHODOLOGY

The main objective of this study is to use statistical models for road accidents at Penang Island. The research hypothesizes that there are some controllable factors that contribute to road accidents such as behavioral preventive measures and driver behavior. Therefore, delineation and subsequent improvement of these factors should improve the road safety in Penang Island. With this aim, the analysis was performed in three essential steps, selection of study area, questionnaire design and analysis, and regression model for road accidents in Penang Island.

-*Collection of the data:* In this study, there were two sets of data, the first one is the primary data which can be collected by the questionnaire and the second set is the secondary data (from Penang Island police station). Different regression models were tested to find the best model that can fit the data. Meanwhile, the interview assisted questionnaire is very important for collecting the primary data. Questionnaire alone might not be sufficient because some of the respondents do not understand the question properly, then their replying will not be correct, furthermore, the interview assisted questionnaire will enable collecting the required information directly.

The questionnaire has been conducted in Klang city. Comparison of the primary and secondary data was achieved. 346 questionnaires have been distributed. Around 37 questionnaire form has been neglected since some of the respondents have left it blank, others have partially filled it. The researcher has interviewed with people, explains the purpose of conducting the questionnaire and made some clarifications regarding any question that has been raised.

The framework for this research is for the comparison between the primary and secondary data and then develops the appropriate model. This study involves the development of model which can be beneficial and reliable by other research work. The scope has commenced from the selection of the study area which is Klang city and the reasons justifying the selection of this study area. The second step involves the design of the questionnaire. The questionnaire has been approved and certified by some experts in the relevant field and from different road departments and institutions like Jabatan Kerja Raya (JKR). The questionnaires have been collected from 309 participants.

The secondary data have been collected from Klang city after getting an authorization letter from Universiti Kebangsaan Malaysia. The secondary data were used to develop the model using Microsoft excel 2007.

Finally, recommendations on how to minimize road accidents have been established. The structural framework for questionnaire study and structural framework for modeling of road accidents were illustrated in Figures 1 and 2 respectively. The most traffic accident analysis in Malaysia was based on a larger geographical area like one or more cities in Malaysia. The focus of this research is to apply the statistical or other model in the city of Klang from a data obtained from Penang Island police station. This can help to better analyses its own accident pattern and take certain corrective actions. The primary objective of the study is to develop a statistical model for the city that identifies the impact of the time on the accident rate.

In statistics, nonlinear regression is a form of regression analysis in which observational data are modeled by a function which is a nonlinear combination of the model parameters and depends on one or more independent variables. The data are fitted by a method of successive approximations. Modeling is a kind of art that you need to trial and error. Each time, the researcher guesses what the model is then he has computed the trend line and the R-squared. Among all guesses, the researcher has decided the best model in which it is produces the highest R-square and tends to explain the data plot. That is why the scattered plot matters.

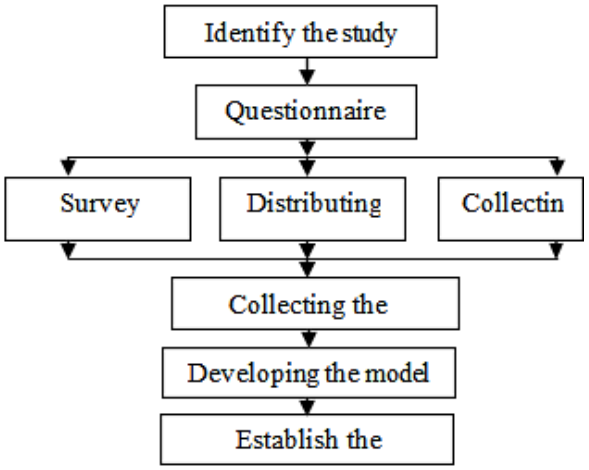

Fig. 1. Research structural framework for questionnaire study

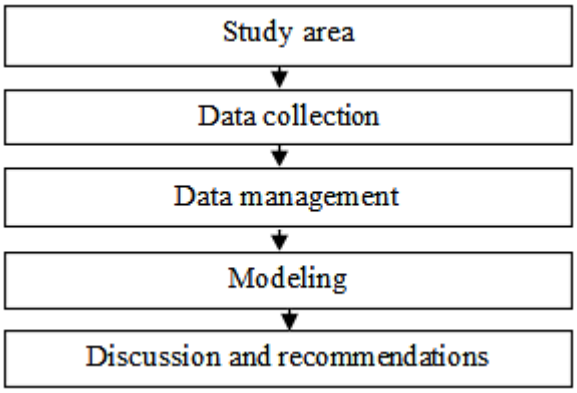

Fig. 2. Pro Research ject structural frameworks for modeling of road accident data in Klang

Modeling of road accidents and time (1996-2010) was done using polynomial regression. Polynomial regression is a form of linear regression in which the relationship between the independent variable x (years) and the dependent variable y (number of accidents) is modeled as an nth order polynomial. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted E (y/x).

III. RESULTS AND DISCUSSIONS

-*Questionnaire Analysis:* the questionnaire that has been conducted throughout this research is carried out in Penang as presented in figure 3. the following section will present each question in the questionnaire and the accompanied analysis.

Q1: Do you find driving a car is enjoyable and rewarding?

TABLE I. DRIVING A CAR IS ENJOYABLE AND REWARDING

Disagree strongly	medium	Strongly agree

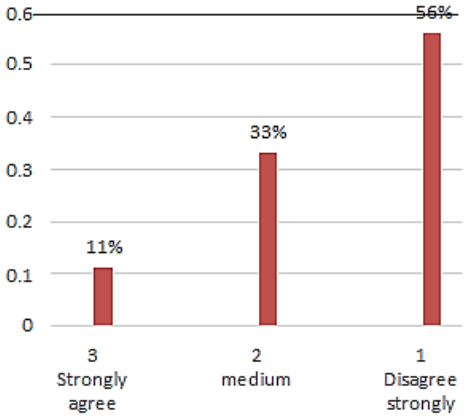

Fig. 3. Driving a car is enjoyable and rewarding

Q2: I perform all appropriate checks before driving or riding e.g. checking the water level, brake oil, clutch oil, the trifurcating lights, using the seat belt.

TABLE II. CAR CHECK PRIOR TO DRIVING

Disagree strongly	medium	Strongly agree

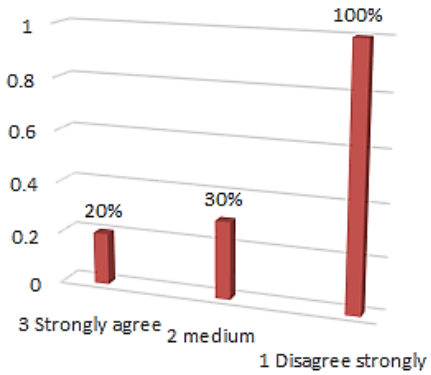

Fig. 4. Car check prior to driving

Q3: When I am driving in interweaving streams of fast moving traffic, with many other Drivers often

TABLE III: CAR CHECK PRIOR TO DRIVING

Disagree strongly	medium	Strongly agree

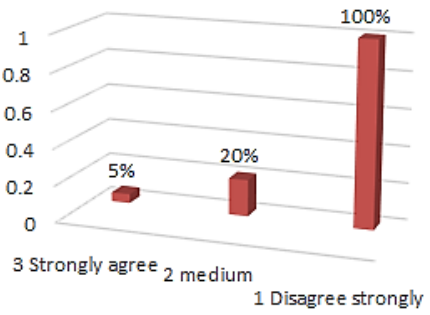

Fig. 5. Car changing lanes

Q4: When a motorcycle collides it is typically the fault of the motorcyclist.

TABLE IV. REASONS FOR MOTORCYCLE COLLIDE

Disagree strongly	NO opinion	Strongly agree

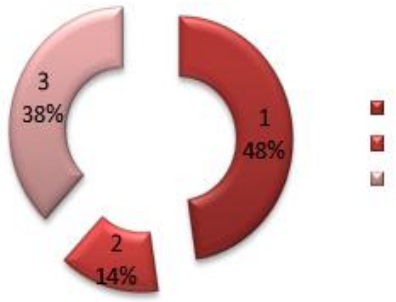

Fig. 6. Reasons for motorcycle collide: 1)strongly agree, 2)no opinion, 3)disagree strongly

Q5: On the open road, you can be suddenly surprised by the appearance of a motorcycle coming from behind you.

TABLE V. PPEARANCE OF MOTORCYCLE COMING FROM BACK

Disagree strongly	NO opinion	Strongly agree

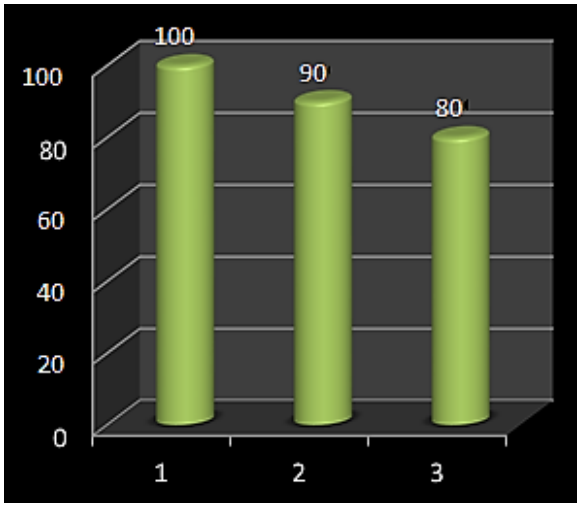

Fig. 7. Appearance of motorcycle coming from the back: 1)strongly agree, 2)no option, 3)disagree strongly

Q6: How often do you become impatient with a slow driver in the outer lane and overtake them on the inside?

TABLE VI. DRIVING PERTINENCE

Never	Very Rarely	Rarely	Sometimes	Often	Nearly	Always

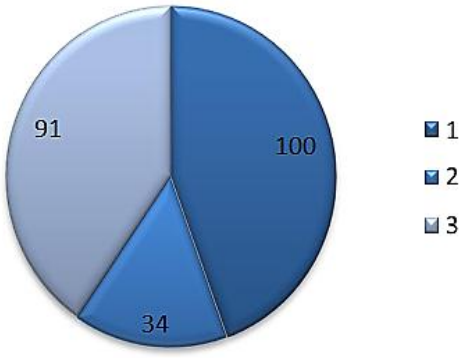

Fig. 8. Driving pertinence: 1)nearly always, 2)sometimes, 3)never

As mentioned earlier, two sets of data were adopted. The first set is the primary data in which questionnaire has been distributed to 225 participants and their feedback were analyzed. Table VII presents the feedback of participants to the questionnaire.

The second set of data was represented by the modeling of road accidents in term of number of fatality versus years. Modeling of road accidents and time was conducted using polynomial regression (Equation 1) as shown in Figure 4. Polynomial regression is a form of linear regression in which the relationship between the independent variable x (years) and the dependent variable y. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted E (y/x).

TABLE VII. QUESTIONNAIRE FEEDBACK

Question No.	% strongly agree	% medium	% disagree strongly
1	56	33	11
2	13	20	67
3	80	16	4
4	48	14	38
5	37	33	30
6	45	15	40
7	72	21	7
8	53	42	5
9	17	33	50
10	39	42	13
11	66	26	8
12	28	14	58
13	48	43	9
14	22	26	52

In statistics, nonlinear regression is a form of regression analysis in which observational data are modeled by a function which is a nonlinear combination of the model parameters and depends on one or more independent variables. The data are fitted by a method of successive approximations. Modeling is a kind of art that you need to do trial and error. Each time, the researcher guessed what the model (in the experiment above, the researcher guessed that it is power curve), then the researcher computed the trend line and the R-squared. Among all of the guesses, the researcher decided the best model is the

model that produces the highest R- square and tend to explain the data plot. This is the reason why the scattered plot matters.

An equation in which one or more terms have a variable of degree 2 or higher is called a nonlinear equation. A nonlinear system of equations contains at least one non-linear equation. Excel spreadsheet provides six possible trends: linear, logarithmic, polynomial, power, exponential and moving average. Using some linear transformation, the researcher may see how this non-linear transformation actually works and then some more nonlinear regression types such as square root and reciprocal curves. The data that have been used for developing the model were illustrated in Table VIII as below.

TABLE VIII. DATA USED FOR DEVELOPING THE MODEL

No.	Year	No. of dead people
1	1996	7935
2	1997	7991
3	1998	7317
4	1999	8252
5	2000	10389
6	2001	11682
7	2002	11838
8	2003	12249
9	2004	13571
10	2005	13619
11	2006	13667
12	2007	13756
13	2008	13512
14	2009	13259
15	2010	13219

The current model reveals that variability in road accident number is only explained by $R^2 = 0.779$ by the years from 1996 to 2010.

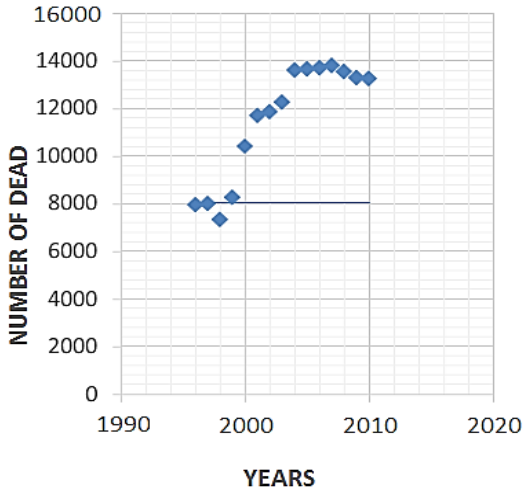

Figure 3: Polynomial nonlinear regression for modeling of fatality based on time

$y = 1E\text{-}303x^{92.946}$ $R^2 = 0.779$
Where,
Y = fatality (number of dead people)
X = Years
R^2 = Coefficient of determination

IV. CONCLUSIONS

The increase of road accidents is in link with the rapid growth in population, economic in development, industrialization and motorization encountered by the country. Since 1970's Malaysia has experienced a remarkable growth in these sectors. In facts, there is an increase in Malaysian population from 10.4 million in 1974 to 26.1 million in 2005 at an average growth rate of about 2.1 % per year. Furthermore, the total length of road had also increased from 11161 km in 1974 to 71814 km in 2005 to accommodate an increase in numbers of vehicles in Malaysia. This also led to an increase of ownership from 9.6 persons per vehicle in 1974 to 1.7 persons per vehicle in 2005. The total numbers of registered vehicles also increased from 1090279 to 15026660 vehicles in 2005 according to [19].

V. ACKNOWLEDGEMENTS

This work was supported by Universiti Kebangsaan Malaysia. The authors are high appreciated that support.

REFERENCES

[1] Y. Dapilah, F. Guba, B. Y., & E. Owusu-Sekyere, "Motorcyclist characteristics and traffic behaviour in urban Northern Ghana: Implications for road traffic accidents," *Journal of Transport & Health*. 2016.

[2] S. Kar, S. C. Das, A. Tiwari, & I. Pharveen, "Pattern of Road Traffic Accidents in Bhubaneswar," Odisha. Clinical Epidemiology and Global Health, 2015.

[3] A. C. Hayley, L. A. Downey, B. Shiferaw, & C. Stough, "Amphetamine-type stimulant use and the risk of injury or death as a result of a road-traffic accident: A systematic review of observational studies," European Neuropsychopharma-cology, vol. 26, no. 6, pp. 901-922, 2016.

[4] P. Batailler, M. Hours, M. Maza, P. Charnay, H. Tardy, C. Tournier, & E. Javouhey, "Health status recovery at one year in children injured in a road accident: A cohort study. Accident Analysis & Prevention," vol. 71, pp. 267-272, 2014.

[5] F. Prieto, E. Gómez-Déniz, & J. M. Sarabia, "Modelling road accident blackspots data with the discrete generalized Pareto distribution," Accident Analysis & Prevention, vol. 71, pp. 38-49, 2014.

[6] Chung, Y., & Chang, I. "How accurate is accident data in road safety research? An application of vehicle black box data regarding pedestrian-to-taxi accidents in Korea," Accident Analysis & Prevention, vol. 84, pp. 1-8, 2015.

[7] A. Azadeh, M. Zarrin, & M. Hamid, "A novel framework for improvement of road accidents considering decision-making styles of drivers in a large metropolitan area," Accident Analysis & Prevention, vol. 87, 17-33, 2016.

[8] B. Dadashova, B. Arenas-Ramírez, J. Mira- McWilliams, & F. Aparicio-Izquierdo, "Methodological development for selection of significant predictors explaining fatal road accidents," Accident Analysis & Prevention, vol. 90, pp. 82- 94, 2016.

[9] A. Theofilatos, G. Yannis, P. Kopelias, & F. Papadimitriou, "Predicting road accidents: a rare-events modeling approach,". *Transportation Research Procedia*, vol. 14, pp. 3399-3405, 2016.

[10] C. Xu, P. Liu, W. Wang, & Z. Li, "Evaluation of the impacts of traffic states on crash risks on freeways," *Accident Analysis & Prevention*, vol. 47, pp. 162-171, 2012.

[11] L. Ramos, L. Silva, M. Santos, &J. Pires, "Detection of road accident accumulation zones with a visual analytics approach," *Procedia Computer Science*, vol. 64, pp.969-976, 2015.

[12] G. Andrienko, N. Andrienko, P. Jankowski, D. Keim, M. Kraak, A. MacEachren, S. Wrobel, "Geovisual analytics for spatial decision support: Setting the research agenda," International Journal of Geographical Information Science, vol. 21, no. 8, pp.839-857, 2007.

[13] J. Kazmi and S. Zubair, "Estimation of vehicle damage cost involved in road traffic accidents in Karachi, Pakistan: a geospatial perspective," *Procedia engineering*, vol. 77, pp.70-78, 2014.

[14] G. Goel, & S. Sachdeva, "Analysis of road accidents on NH-1 between RD 98km to 148km," Perspectives in Science, 2016.
[15] Ministry of Transport. Road Safety Plan of Malaysia 2006-2010. Kuala Lumpur: Road Safety, 2006.
[16] R. Lyons, H. Ward, H. Brunt, S. Macey, R. Thoreau, O. Bodger, & M. Woodford, "Using multiple datasets to understand trends in serious road traffic casualties," Accident Analysis & Prevention, vol. 40, no.4, pp. 1406-1410, 2008.
[17] G. Jacobs and P Hutcison. "A study of accident rates in Developkg Countries," *Department of the Environment Department of Transport, TRRL Laboratory Report LR 546.* Crowthorne, 1973, (Transport and Road Research Laboratory).
[18] R .Jorgensen, Cost and safety effectiveness of highway design elements. *NCHRP Report*, 1978.
[19] M. Mustafa, Overview of current road safety situation in Malaysia. Universiti Putra Malaysia, Department of Statistics, 2005.
[20] U. Radin, Updates of road safety status in Malaysia. IATSS research, vol. 1, no. 29, pp.106-108, 2005.