

The Influence Factors of Highway Traffic Accident and Accident Rates Model

Ruize Ma

School of arts and sciences, Rutgers University, New Brunswick campus, NJ 08901, US ruizema1@gmail.com

Keywords: automobile, highway accident rates, multiple regression.

Abstract. Road traffic accidents are traffic anomalies that occur under the interactions of multiple factors such as people, cars, roads and the environment, the road conditions and traffic regulations are the basic elements of the traffic, which cannot be neglected in the accident process, especially those accidents happen in highways. Thus, it will be of great significance to study the relationship between the traffic accidents and the factors that are under the control of the highway department. Based on this, the goal of this study was to explore the impact of design variables (such as number of access points per mile, speed limit, number of signalized interchanges per mile, width in feet of outer shoulder on the roadway and so on) that are under the control of the highway department, on accident rates.

1. Introduction

Traffic safety is an important part of road construction and management, it relates to the life safety and property safety of citizens, and it is also an important condition for achieving economic benefits of highway construction. Among the three elements of traffic accidents -- people, cars and the environment, people are the key factors of traffic accidents, including the drivers, passengers and pedestrians; the mechanical properties of the automobile itself are the premises of ensuring driving safety; environmental factors sometimes are the fuses of accidents, environmental factors not only include the integrity of road transport facilities, such as line type, signs and the integrity of the road, but also include the size of traffic flow, vehicle composition and natural environmental factors, such as season and weather.

From the investigation results of traffic accidents, accidents that are directly caused by poor road conditions are rare, that is, in accident cases, traffic management departments scarcely regard the road conditions as the direct cause of those accidents. The occurrence of this phenomenon is largely due to the lack of awareness of the road conditions, responsible person put too much emphasis on the reasons of people and vehicles. The occurrence of road traffic accidents usually is the result of the combined action of many factors, such as people, vehicles and roads. However, traffic management departments often regard one factors as the main factor in the accident investigation site. Thus, in most cases, the road conditions are identified as secondary causes when determining the cause of the accident, even ignored. However, road conditions are closely related to the traffic accidents. Snow and wet road, road without central separation zone, lack of sight distance, too long straight line, too high slope or improper connection of straight line and curve, may all be the causes of traffic accidents.

Past research on modeling accident rate has been diverse, both empirically and methodologically. Based on the data of Korea Expressway, Chang et al. did an empirical study and found that in the basic section, tunnel section and toll station areas of highway, there was an approximate U-shaped curve relationship between accident rates and the degree of traffic load (the ratio of flow rate to road capacity) An; when the traffic flow state is the same, the accident rate of the toll station areas is significantly greater than that of other road [1]. Hikosaka analyzed the traffic accidents occurred in the basic sections of some highways in Japan, and found for unsaturated flow, there was a U-shaped curve relationship between the accident rate and the degree of traffic load; and for saturated flow, the accident rate will increase rapidly with the increase of the traffic density [2]. Using hierarchical tree regression method, Matthew et al. quantified the impact of road characteristics on accident rates, research result showed that the road line shape and road conditions are the most important factors in traffic accidents [3]. Other run-off-roadway accident studies have examined particular roadside features such as roadway guardrail systems, utility poles, bridges, sign supports, side slopes and ditches and fences and their effect on the frequency and severity of accidents [4-6].

When establishing the prediction model of traffic accidents, we should describe and analyze it based on the road conditions, because it is quite difficult to quantitatively describe human factors and vehicle factors, which have a necessary link with road conditions. Based on this, the goal of this analysis was to explore the impact of design variables (such as number of access points per mile, speed limit in 1973, number of signalized interchanges per mile, and width in feet of outer shoulder on the roadway) that are under the control of the highway department, on accident rates.

2. Data

The data comes from an unpublished master's paper by Carl Hofstede. They relate the automobile accident rate, in accidents per million vehicle miles to several potential terms. The data include 39 sections of large highways in the state of Minnesota in 1973, the attributes of each section include ADT (average daily traffic count in thousands), Trks (truck volume as a percent of the total volume), Lane (total number of lanes of traffic), Acpt (number of access points per mile), Sigs (number of signalized interchanges per mile), Itg (number of freeway-type interchanges per mile), Slim (speed limit in 1973), Len (length of the highway segment in miles), Lwid (lane width, in feet), Shld (width in feet of outer shoulder on the roadway), Hwy (an indicator of the type of roadway or the source of funding for the road; 0 if MC, 1 if FAI, 2 if PA, 3 if MA) and Rate(1973 accident rate per million vehicle miles).

3. Method and Empirical Results

The goal of this analysis was to understand the impact of design variables, Acpt, Slim, Sig, and Shld that are under the control of the highway department, on accidents. So we made Rate (1973 accident rate per million vehicle miles) as the dependent variable, and ADT, Trks, Lane, Len, Lwid, Shld as independent variables to conduct multiple regression. The multiple regression model can be expressed as follow:

 $y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\ldots+\beta_kx_k+u$

Where \mathcal{Y} is the dependent variable, x_i (i = 1, 2, ..., k) are the explanatory variables, β_i (i = 1, 2, ..., k) are regression coefficients, u is the stochastic error term.

Table 1 the empirical results of multiple regression Estimate Std.Error t value Pr(>|t|)2.984 0.00541** (Intercept) 20.61057 6.906518 Lwid -0.915189 0.554814 -1.65 0.10882 0.00171** Shld -0.3127190.091317 -3.425 Lane 0.102729 0.322249 0.319 0.75196 Len -0.103358 0.038955 -2.653 0.01231* ADT 0.002048 0.023515 0.087 0.93115 Trks -0.281014 0.119066 -2.36 0.02453* Multiple R-squared: 0.5383, Adjusted R-squared: 0.4517 F-statistic: 6.218 on 6 and 32 DF, p-value: 0.0002089

The results of the regression analysis using all 13 explanatory variables are given in Table 1.



As is shown in Table 1, these 6 explanatory variables account for almost 53.85% of the highway data. The β coefficients of all the variables are relatively small, except for length of the highway segment in miles, width in feet of outer shoulder on the roadway (1962), and truck volume as a percent of the total volume. As F-statistic is highly significant, and which provides evidence for the presence of a linear relationship between all 6 variables and the response variable, the β coefficients of width in feet of outer shoulder on the roadway (1962) have significant t ratios. The discrepancy is explained by the next analysis which is a correlation matrix between all the 7 variables, the result is shown in Table 2.

	Rate	Lwid	Shld	Lane	Len	ADT	Trks
Rate	1.000000	-0.005619	-0.386907	-0.032979	-0.465290	-0.028570	-0.512522
Lwid	-0.005619	1.000000	-0.042896	0.095723	-0.310650	0.127878	-0.155271
Shld	-0.386907	-0.042896	1.000000	0.481771	-0.104926	0.457307	0.006135
Lane	-0.032979	0.095723	0.481771	1.000000	-0.202504	0.823930	-0.153324
Len	-0.465290	-0.310650	-0.104926	-0.202504	1.000000	-0.271569	0.495943
ADT	-0.028570	0.127878	0.457307	0.823930	-0.271569	1.000000	-0.096682
Trks	-0.512522	-0.155271	0.006135	-0.153324	0.495943	-0.096682	1.000000

Table 2 the correlation matrix between variables



Figure 1 Scatter plot of average daily traffic count in thousands and total number of lanes of traffic Table 3 the empirical results of the new multiple regression

	Estimate	StdcError	t-value	Pr(> t)					
(Intercept)	20.50753	6.69974	3.061	0.004289**					
Lwid	-0.88476	0.53893	-1.642	0.109873					
Shld	-0.28521	0.07751 -3.68		0.000803***					
Len	-0.10579	0.03698	-2.86	0.007185**					
Trks	-0.28709	0.11412	-2.516	0.016765*					
Multiple R-squared: 0.5329, Adjusted R-squared: 0.4779									
F-statistic: 9.697 on 4 and 34 DF, p-value: 2.413e-05									



As is shown in Table 2, the correlation matrix above shows the correlation coefficient (Pearson's r) for each of the explanatory variable with every other variable. Average daily traffic count in thousands and total number of lanes of traffic are also closely correlated with each other at 0.823929. Some variables are more closely correlated with truck volume as a percent of the total volume than others, and truck volume as a percent of the total volume show the highest correlation coefficients for 1973 accident rate per million vehicle miles at -0.512522. And the two variables are negative related, which means if truck volume is bigger, 1973 accident rate per million vehicle miles are lower.

Explanatory variables are closely correlated with truck volume as a percent of the total volume. This is because β coefficients are partial correlation coefficients, and each β describes the relationship between the corresponding X and the response Y taking into account the other X variables. In other words, each of 10 β coefficients tells us about its respective variable's contribution after allowing for the contributions of other explanatory variables. Therefore, if a variable is highly correlated with another it will have no additional contribution to make over and above the contribution of the other. In this particular data set, width in feet of outer shoulder on the average daily traffic count in thousands and total number of lanes of traffic are closely correlated. Each tells only part of the story and we cannot reliably estimate their effects. Unreliable parameter estimate means large standard error and lower t statistic. How closely the two variables are correlated can be judged from Figure 1.

Regression Analysis with several explanatory variables tells us about the effect of each individual variable in the model taking into account the effects of other variables included in the model. The magnitude, significance, and interpretation of any one β coefficient depend on what other variables are included. A method of selecting a sub-set from a list of explanatory variables is Stepwise Regression Analysis. Most computer software have a facility for carrying out stepwise forward selection and backward elimination of variables based on their overall contribution to the F statistic. Significance of F can be set by the individual or left to the default value.

Stepwise regression removes and adds variables to the regression model for the purpose of identifying a useful subset of the predictors. Stepwise first finds the explanatory variable with the highest correlation R^2 to start with. It then tries each of the remaining explanatory variables until it finds the two with the highest R^2 . Then it tries all of them again until it finds the three variables with the highest R^2 , and so on. The overall R^2 gets larger as more variables are added. In conclusion, stepwise regression is useful in the early exploratory phase of analysis, but not to be relied on for the confirmatory stage. Once the exploratory stage has identified the significant explanatory variables they can then be further assessed.

In resorting to the stepwise procedure, the following should be done in mind:

1). By chance about 5% of sample relationships can be expected to be significant, and one can wrongly single them out as being truly significant.

2). When there are strong relationships between several explanatory variables, stepwise procedures tend to exclude one or more of these variables. One may then mistakenly conclude that the excluded variable is unimportant.

3). If two variables X_1 and X_2 are positively related but their effects on Y have opposite signs, or X_1 and X_2 are negatively related but their effects on Y have the same signs, stepwise procedures may exclude one or both. We then understate the importance of both variables.

4). Stepwise regression tends to choose only those observations which have no missing values for any of the variables. So the final model obtained by the Stepwise procedure will have fewer subjects if the data set has missing values. Hence the model selected by Stepwise may not fit a larger subsequent data set well.

5). the p-values are to be discounted because they do not take into account the very many tests that have been carried out.

6). Different methods of stepwise like 'forward selection' and 'backward elimination' are likely to produce different models. Also the same model seldom emerges when a second data set is analyzed.

Through conducting Stepwise Regression Analysis, we performed a new multiple regression analysis excluding ADT (average daily traffic count in thousands) and Lane (total number of lanes of traffic). The final empirical results are shown in Table 3

4. Conclusion

The impact that traffic accidents have on society is significant. There are approximately thousands of people are killed and thousands more injured by traffic accidents all over the world. The main inducing factor of highway traffic accidents is people, but it is undeniable that there are still a large number of traffic accidents caused by direct or indirect road conditions. It is difficult to collect dynamic information regarding drivers and cars when traffic accidents occur, so this study explore the impact of highway load conditions on accident rates. The results obtained here, by exploring a broad range of variables including highway geometry, traffic and environmental characteristics, provide valuable insight into the underlying relationship between risk factors and vehicle accidents.

The findings of this study are suggestive but limited in that they are based on accidents in the large highways in the state of Minnesota in 1973. Due to limited data, we are unable to study accident frequencies for regional differences.

References

- [1]. Chang, J., Oh, C., & Chang, M. (2000, June). Effects of traffic condition (v/c) on safety at freeway facility sections. In Transportation Research, E-Circular, Fourth International Symposium on Highway Capacity Proceedings (pp. 200-208).
- [2]. HIKOSAKA, T. (2001). A Statistical Analysis of Relation-ship Between Traffic Flow Condition and Accident Rate on Basic Motorway Section(Doctoral dissertation, Nagoya: Nago-ya University).
- [3]. Karlaftis, M. G., & Golias, I. (2002). Effects of road geometry and traffic volumes on rural roadway accident rates. Accident Analysis & Prevention, 34(3), 357-365.
- [4]. Bateman, M., Howard, I., Johnson, A., & Walton, J. (1998). Model of the performance of a roadway safety fence and its use for design. Transportation Research Record: Journal of the Transportation Research Board, (1647), 122-129.
- [5]. Reid, J., Sicking, D., Faller, R., & Pfeifer, B. (1997). Development of a new guardrail system. Transportation Research Record: Journal of the Transportation Research Board, (1599), 72-80.
- [6]. Ray, M. H. (1999). Impact conditions in side-impact collisions with fixed roadside objects. Accident Analysis & Prevention, 31(1), 21-30.