International Journal of Applied Science and Engineering
2009. 7, 2: 169-175

Application of Prediction Techniques to Road Safety in Developing Countries

Dr. Jamal Al-Matawah * and Prof. Khair Jadaan

Department of Civil Engineering: College of Technological Studies in Kuwait and the University of Jorda

Abstract: The dramatic increase in vehicle travel in developing countries calls for the effective introduction of features that reduce traffic accidents. An important piece of information for such an introduction lies in the prediction of accidents and their fatalities, which is addressed in this paper. Smeed's model was originally developed for the prediction of traffic fatalities in both developed and developing countries. More reliable prediction models are developed for a number of Arab countries, producing a much less absolute percentage error than those of Smeed's model. Regression analysis was applied to time-series data in the studied countries, producing an absolute percentage error as low as 7.67 for Saudi Arabia and 12.17 for Kuwait. An accident prediction model that relates accident frequency in Kuwait to various contributory factors is developed using the Generalized Linear Modelling (GLM) technique. The final model shows that age, nationality, aggressive driver behaviour, dangerous offences, perception of effectiveness of enforcement, marital status, speed, and experience are the main contributory factors that lead to accident involvement.

Keywords: road safety; prediction models; developing countries.

1. Introduction

Road traffic accidents and their resulting fatalities may be regarded as a growing social and economic problem, especially in developing countries where the resources are limited. The World Health Organization has predicted that traffic fatalities will be the third leading cause of death worldwide by 2020 [1]. The effects of some of the contributing factors to traffic fatalities have been studied and relationships for predicting these fatalities have been developed by [2-8]. Yet, these relationships produced somewhat large deviations between the expected and the observed fatalities. These deviations were greatest in developing countries and the need arises for a more realistic relationship to predict road traffic fatalities with greater accuracy. In addition, these relationships failed to incorporate many significant contributory factors.

Attempts to produce prediction models for traffic fatalities avoiding the above-mentioned pitfalls are discussed in this paper. The study uses a regression analysis of time-series fatality data for the development and testing of the model for the statistics available from the UAE, Jordan and Qatar. The Generalized linear model (GLM) technique is also used to develop a model that incorporates various significant contributory factors.

* Corresponding author; e-mail: jamaln1@hotmail.com
Accepted for Publication: June 18, 2009

2. Prediction Models

The most significant early attempt to predict road traffic fatalities was that of Smeed [9] who used data for road fatalities, vehicles and population for 1938 from 20 countries, the majority of which were European, to derive a rather more complicated relationship known as Smeed's Law, which is expressed by the following formula.

for

$$D=0.0003(VP^2)^{1/3} \quad (1)$$

or

$$D/V= 0.0003 (V/P)^{0.67} \quad (2)$$

$$D/P= 0.0003 (V/P)^{0.33} \quad (3)$$

Where

D= number of road fatalities
V= number of vehicles
P= population

Smeed's formula was tested on a number of occasions and found to produce a large percentage deviation of the expected road fatalities from those observed, that reached 191 percent. The validity of Smeed's formula with data from developing countries was tested on a number of occasions. [10] using accident data in Kuwait for the period between 1969 and 1980, found that the average absolute percentage error for the estimated fatalities reached 30.85. Using data from Saudi Arabia for the fourteen year period between 1974 and 1987, Smeed's formula produced an average absolute percentage error of 35.8 [11].

The relatively big deviation between the observed and expected values dictated the need to develop more reliable models than Smeed's formula. A number of attempts were made, producing the following results:

a. Applying the least square method to 20-year (1969-1989) data of Kuwait developed the following regressed equation [2].

$$D/V= 1288.25 (V)^{1.138} (P)^{-1.138} \quad (4)$$

The average absolute percentage error was found to be only 12.17, much less than that of Smeed.

b. Time-series data between 1974 and 1987 for Saudi Arabia were used for model development [11]. The data were fitted successfully to a logarithmic model of the following form.

$$D= \text{In} (V^{849.3} P^{-608.1}) \quad (5)$$

(R^2=0.85)

The average absolute percentage error was found to be only 7.67.

3. Further development of prediction models.

Smeed's law is usually criticized for having the number of vehicles (V) on both sides of the equation and that there is a considerable deviation between the expected and actual number of road fatalities. Therefore, the need arises to develop a predictive model of road fatalities which fits the data for developing countries and provides better estimates than Smeed's equation.

Time-series data of road fatalities vehicles and the population for the UAE, Jordan and Qatar between 1990 and 2004, shown in Tables 1, 2, and 3 respectively, were used for model development.

Table 1. Fatalities in the UAE according to Smeed's equation and regression analysis from 1990 to 2004

Year	Vehicles	Popula-tion	Fatalities	Fatalities Estimates By Regression	Fatalities Estimates By Smeed's Equation	% of error by regres-sion	% of error by Smeed's equation
1990	303284	1844300	394	450	303	14.3	-23.1
1999	309539	1908800	490	454	312	-7.3	-36.3
1992	344850	2011400	510	486	335	-4.7	-34.3
1993	398788	2083100	567	539	360	-5.0	-36.5
1994	428149	2230000	600	563	386	-6.2	-35.7
1995	437945	2377453	563	568	406	0.8	-27.9
1996	442700	2477899	492	569	419	15.6	-14.9
1997	468440	2624000	619	590	443	-4.7	-28.4
1998	539407	2759000	646	657	480	1.7	-25.6
1999	557668	2938000	624	669	507	7.2	-18.8
2000	575929	3108000	673	682	532	1.3	-21.0
2001	605696	3290000	733	705	561	-3.8	-23.4
2002	633817	3754000	743	717	622	-3.5	-16.2
2003	661937	4041000	777	756	736	-2.7	-5.3
2004	690058	4320000	729	754	703	3.5	-3.5

Table 2. Fatalities in Jordan according to Smeed's equation and regression analysis from 1990 to 2004

Year	Vehicles	Population	Fatalities	Fatalities Estimates By Regression	Fatalities Estimates By Smeed's Equation	% of error by regres-sion	% of error by Smeed's equation
1990	254617	3453000	379	345	434	-8.9	14.5
1999	259196	3888000	379	398	473	5.0	24.9
1992	276301	4012000	388	426	493	9.9	27.0
1993	291347	4152000	440	453	514	2.9	16.8
1994	304893	4200000	443	469	526	5.9	18.7
1995	321373	4290100	469	492	543	4.9	15.8
1996	342337	4444000	552	525	567	-4.9	2.7
1997	362811	4600000	577	559	592	-3.1	2.6
1998	389196	4755800	612	596	619	-2.6	1.1
1999	418433	4900000	676	635	647	-6.1	-4.3
2000	473339	5039000	686	693	687	1.0	0.15
2001	509832	5182000	783	737	718	5.8-	-8.3
2002	542812	5329000	758	779	747	2.7	-1.5
2003	571498	5480000	832	818	774	-1.7	-6.9
2004	612330	5650000	818	868	808	6.1	-1.2

Table 3. Fatalities in Qatar according to Smeed's equation and regression analysis from 1990 to 2006

Year	Vehicles	Population	Fatalities	Fatalities Estimates By Regression	Fatalities Estimates By Smeed's Equation	% of error by regres-sion	% of error by Smeed's equation
1990	161262	422145	96	63	92	-34.4	-4.2
1999	177082	345658	96	70	97	-27.1	1.0
1992	190050	449606	116	76	101	-34.5	-12.9
1993	203001	464009	84	81	106	-3.6	26.2
1994	207912	467402	52	84	108	61.5	108.0
1995	217802	494225	99	88	113	-11.1	14.0
1996	231006	510070	89	94	118	5.6	32.0
1997	249787	526429	96	102	123	6.3	28.3
1998	269510	543315	106	111	129	4.7	21.7
1999	284018	560746	96	118	134	22.9	39.7
2000	299611	578470	85	124	139	45.9	64.0
2001	319318	595321	110	133	145	20.9	31.9
2002	348840	616151	114	146	153	28.1	34.1
2003	366532	717984	150	154	172	2.7	14.8
2004	402006	755163	164	170	184	3.7	12.2
2005	457239	796186	206	181	199	-12.1	-3.4
2006	544013	837209	270	233	218	-13.7	19.3

The inspection of the data reveals that it can be described as unstable and non-homogeneous, which led to the idea of using a natural logarithm function of the form:

$$D= cVa\ Pb \quad (6)$$

The parameters a, b and c are estimated by first converting equation (6) into a linear form:

$$\ln(D)=\ln c+a\ln V+b\ln P \quad (7)$$

The parameters a, b and c were estimated from the least square regression analysis, resulting in the following relationship

$$\ln(D)=e^{1311.872}+384.502\ln V-420.486\ln P \quad (8)$$

The equation may by written in terms of the untransformed variables by exponentiating the log-transformed terms:

$$D=e^{1311.872} * V^{384.502} * P^{-420.486} \quad (9)$$

The statistical significance of a regression coefficient is tested by advancing the null hypothesis which asserts that the magnitude of a partial regression coefficient is drawn from a statistical distribution with a mean value of 0. The null hypothesis was tested by the student's t and the results indicated that both partial regression coefficients are statistically significant at the 5% level of significance.

Comparing the actual and predicted number of fatalities, it was found that all models gave better estimates than Smeed's equation. The average absolute percentage errors were 23.4% and 27.5% for Smeed's equation for the UAE and Qatar respectively. However, for Jordan, whose population is greater than that of the UAE and Qatar, the percentage errors were only 9.7%. These compare with the regression percentages of 5.5, 20.0 and

4.8 for the respective countries, as shown in Table 4.

Tables 1, 2 and 3 show the predicted fatalities by Smeed's equation and the regression model respectively for each year for the period covered by the model and for each of the studied countries. The tables also include the actual number of fatalities for comparison. It is unsurprising that the models produce predicted fatalities that correlate significantly with the actual data since they use the same data for prediction as were used to generate their parameters.

Table 4. Summary of average percentage error of Smeed's equation compared with regression equation in the UAE, Jordan and Qatar

Country	Smeed's equation	Regression equation
UAE	23.4	5.5
Jordan	9.7	4.8
State of Qatar	27.5	20.0

4. Development of an accident prediction model using the GLM technique

The technique of Generalized Linear Modelling (GLMs) was first used by [13] to develop an accident prediction model that related accident frequency to mileage and other relevant variables.

The same technique was applied to accident data about Kuwait, collected using a pre-designed survey questionnaire. The model used the number of accidents a driver had been involved in over the last 10 years as the dependent variable (A) and the number of independent variables believed to have a significant effect on accidents in Kuwait, including age, sex, nationality, education level, marital status, aggressive driving behaviour score, driver education, driver training, usual speed on motorways, years of driving experience, and effectiveness of enforcement.

The backward elimination procedure was used in this research for variable selection and to determine the 'best' fitting model in terms of both the statistical significance (p value < 0.05) and appropriateness of the variables included, using Generalised Linear Modelling with the distribution of the response variable "Poisson distributed".

The procedure starts with all the independent variables and tests them individually for statistical significance, dropping those that are insignificant (eliminating the weakest predictor of the dependent variable). The procedure is better able to handle the multicollinearity of the data (i.e. two or more independents are at a high degree of correlation) than the forward procedure [14].

Stata.9 software was used to run the models and the developed Generalised Linear Model was found to have the following form:

$$\mathbf{A} = 0.198\ (\mathbf{Exposure})^{0.135} \cdot e^{-0.0246\mathbf{B} + 0.31\mathbf{C} - 0.433\mathbf{D} + 0.18\mathbf{E} + 0.06\mathbf{F} - 0.161\mathbf{G} - 0.0386\mathbf{H} + 0.204\mathbf{I}}$$

The computed scale parameter is close to one (unity), indicating that there is little evidence of over-dispersion. Hence, there is little need to adjust the standard errors. The standard errors can be corrected for over-dispersion by multiplying by the square root of the scale parameter. This will change the 'z' value and 'p' value, as illustrated in Table 5.

Table 5 The final accident prediction model after adjusting the standard error for over-dispersion Dependent variable: accidents

Independent variables	Coefficients	Standard Errors	Z value	P value
Age (in years)	-0.0246	0.0089	-2.77	0.006
Nationality (0 if non-Kuwaiti, 1 if Kuwaiti)	0.3088	0.0825	3.74	0.000
Marital status (0 if single, 1 if married)	-0.4330	0.0800	-5.41	0.000
Effectiveness of Enforcement (0 if yes, 1 if no)	0.2042	0.0644	3.17	0.002
Speed (0 if within the speed limit, 1 if exceeding the speed limit)	0.1803	0.0919	1.96	0.050
Number of dangerous offences per year	0.0601	0.0147	4.09	0.000
Aggressiveness (ranging from 1 =always to 5 =never)	-0.1615	0.0593	-2.72	0.007
Experience (in years)	-0.0385	0.0091	-4.21	0.000
Exposure (in Kilometres)	0.1348	0.0498	2.71	0.007
Constant	-1.6210	0.6013	-2.70	0.007
No. of obs. = 1016 Residual df = 1006 Scale parameter = 1 Deviance = 1290.740556 (1/df) Deviance =1.283042 Pearson = 497.685905 (1/df) Pearson = 1.488753 Variance function: V(u) = u [Poisson] Link function : g(u) = ln(u) [Log]				

The resulting identified effects of the contributory factors are summarized as follows:

a. Age (B): the younger the age, the greater the number of accidents.
b. Nationality (C): Kuwaitis were involved in more accidents than were non-Kuwaitis.
c. Marital status (D): single drivers were involved in more accidents than married drivers.
d. Speed (E): speeding on motorways leads to more accidents.
e. Number of dangerous offences per year (F): the greater the number of dangerous offences per year, the higher the number of accidents.
f. Aggressive driver behaviour score (G): the more aggressive the driving, the greater the number of accidents.
g. Experience (H): the more experienced the driver in terms of number of years driven, the less involvement they have in accidents.
h. Effectiveness of enforcement (I): drivers who think that enforcement is ineffective experience a greater number of accidents compared to drivers who perceive enforcement to be effective.

5. Conclusion

Smeed's equation for predicting road fatalities produces a considerable deviation between the expected and actual road fatalities, especially when applied to developing countries, and does not consider the various factors contributing to road accidents. Therefore, Natural Logarithmic models were developed and found to produce more accurate predictions than Smeed's model for selected developing countries. In addition, a GLM was developed for Kuwait to identify the most significant accident contributory factors. These were found to be age, nationality, aggressive driver behaviour, dangerous offences, perception of effectiveness of enforcement, marital status, speed, and experience.

References

[1] Murray and Lopez, 1996. In: C. Murray and A. Lopez, Editors, *"The Global Burden of Disease"*, Harvard Press, Cambridge, MA.

[2] Haight FA. 1980. Traffic safely in developing countries. *Journal of Safety Research*; 12: 50-58.

[3] Jacobs, G., Aeron-Thomas, A., Astrop, A., 2000. Estimating Global Road Fatalities. TRL Report 445. Transporting Research Laboratory, London, England.

[4] Bener A, Ofsu J. B. 1991. Road traffic fatalities in Saudi Arabia. *J In Assoc. Traffic and safety Science*; 15:35-38.

[5] Elvik R. 1995 Analysis of official economic valuations of traffic accident fatalities in 20 motorized countries. *Accid Anal Prev*. 27:237-47.

[6] Haight FA. 1980. Traffic safely in developing countries. *Journal of Safety Research*; 12: 50-58.

[7] Bishai D, Quresh A, James P, Ghaffar A. 2006. National road casualties and economic development. *Health Econ.*; 15:65-81.

[8] Kopits, E., 2005. Cropper, M, Traffic fatalities and economic growth. *Accid Anal prev*. 37: 169-78.

[9] Smeed, R. J. 1949. Some statistical aspects of road safety research. *Journal of Royal Statistical Society Series: A (Ststistics in Society)*; 12, 1: 1-23.

[10] Jadaan, K. S., 1982. A study of accident rates in Kuwait. *Journal of the University of Kuwait (science)* 9: 41-50.

[11] Jadaan, KS., Alenezi, F., AlZahrani, 1992. A. Road fatalities in Saudi Arabia; Trends and prediction, *proceeding of the REAAA workshop*, 3: 19-24.

[12] Jadaan K S., Khalil R., Bener A. 1991. A mathematical model using convex combination for the prediction of the road traffic deaths. *Journal of Computing and information* 2: 139-157.

[13] McCullagh, P. and Nelder, J. A. 1983. Generalized Linear Model. London: Chapman and Hall.

[14] Chatterjee, S., Hadi, A. S.,Price, B. 2000., "*Regression analysis by example"*, New York, Chichester: Wiley and Sons.