Development and Implementation of an Empirical Model on Road Traffic Accidents.

Lukman Abiodun Nafiu, Ph.D.[1*]; David Ibitayo Lanlege, M.Tech.[2]; and Danladi Hakimi, Ph.D.[1]

[1]Department of Mathematics and Statistics, Federal University of Technology, Minna, Nigeria.
[2]Ibrahim Badamasi Babangida University, Lapai, Nigeria.

E-mail: firdaousmama@gmail.com*

ABSTRACT

Road traffic accidents have been a major problem in Nigeria. Due to the increase in population and road users, there have been some short comings in the management of the road traffic accidents by the government. This paper gives a clear picture of the causes of these roads accidents in Nigeria and then comes up with a model that predicts the number of lives that could be lost knowing figures for the fatal, serious, and minor accidents. Regression analysis was used to develop the model and the model results were used to recommend measures that could be taken by the government and the Federal Road Safety Corps in order to reduce the accident counts.

(Keywords: accident, model, road, safety, death, lives, regression analysis)

INTRODUCTION

The various forms of transportation in use today include road, air, water, and rail transportation systems for passengers, flammable and chemical product transportation, and cargo movement (Gustavsson, 1969; Nicholson, 1985). Road transportation is common and it this mode that has given the greatest support to industries for better improvement of the economy of a nation (Nicholson and Wong, 1993).

Traffic accidents in this mode of transportation (road) are the subject of this paper. It dwells more on the measure of controlling road traffic accidents on the Nigerian highways by the Federal Road Safety Commission.

Both experiences and studies have shown that the road traffic accidents in Nigeria have been partially controlled (FRSC, 1988). Based on the statistics, it has been noted that, highways are suffering from serious traffic congestion, poor maintenance, and poor road networks within the cities which are some of the contributing factors to increase in road traffic accidents (Andrew, 2003; Ader, 2008).

However, it is obvious to see from dailies that accidental deaths and casualties of people of different age distribution are continuously published. The net result of this dangerous trend would be a steady brain drain and the scaring effect it has on anticipated foreign investors.

The traffic situation before the establishment of the Federal Road Safety Commission in Nigeria could best be described as chaotic, unpredictable, and indeed dangerous as it was characterized by unprecedented wave of road traffic accidents with attendant colossal human and material losses. Within this era, public awareness and interest in Road Safety was minimal. There was uncoordinated and haphazard licensing drivers and vehicles as well as absence was lacking. Quantitatively, road traffic accidents fatality index as at 1987 was 302 at death per 1,000 vehicles (Daily Trust, 2010).

Several efforts are being made by the Federal Road Safety Commission and by other traffic organizations including Federal Ministry of works and Housing (FMW&H) and the vehicle Investigation Officers (VIO) to control the rate of road accidents, reduce congestion as well as enforce the rules and regulations on the road users.

The term regression was coined by Francis Galton in the nineteenth century to describe a biological phenomenon. The phenomenon that the heights of descendants of tall ancestors tend to regress down towards a normal average (a phenomenon is also known as regression toward the mean). For Galton (1963), regression had only his biological meaning but his work was later extended by Yule and Karl to a more general statistical context (Spiegel and Boxes, 1972). In their work, the joint distribution of the response and explanatory variables is assumed to be Gaussian.

This assumption was weakened by Montgomery (1994). He assumed that the conditional distribution of the response variable is Gaussian, but the joint distribution need not be. In this respect, his assumption is closer to Gauss's formulation.

Regression methods continue to be an area of active research. In recent decades, new methods have been developed for robust regression involving correlated responses such as time series and growth curves, regression in which the predictor variables are curves, images, graphs, or other complex data objects. Regression methods accommodates various types of missing data, nonparametric regression, Bayesians methods for regression, regression in which the predictor variables are measured with error, regression with more predictor variables than observations and casual inference with regression (Griffith, 2009).

In general, regression analysis includes any techniques for modeling and analyzing several variables when the focus is on the relationship between a dependent variable and one or more independent variable. More specifically, regression analysis helps us to understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are fixed (Lindman, 1974).

According to Devore and Peck (1986), regression analysis estimates the conditional expectation of the dependent variable given the independent variables that is, the average value of the dependent variable when the independent are fixed. The focus is on a quartile, or other location parameter of the conditional distribution of the dependent variable given the independent variables called the regression function.

In regression analysis, it is also of interest to characterize the variation of the dependent variable around the regression function, which can be described by a probability distribution. Regression analysis is widely used for prediction and forecasting where its use has substantial overlap with the field of machine learning. Regression analysis is also used to understand which among the independent variables are related to the dependent variable, and to explore the forms of these relationships. In restricted circumstances, regression analysis can be used to infer casual relationships between the independent and dependent variables (Mason, 1992).

A large body of techniques for carrying out regression analysis has been developed. Familiar methods such as linear regression and ordinary least squares regression are parametric, in that the regression function is defined in terms of a finite number of unknown parameters that are estimated from the data. Nonparametric regression refers to techniques that allow the regression function to lie in a specified set of functions, which may be infinite-dimensional (Sridhar, 2003).

The performance of regression analysis methods in practice depends on the form of the data generating process, and how it relates to the regression approach being used. Since the true form of data-generating process is not known, regression analysis depends to some extent on making assumptions about this process. These assumptions are sometimes (but not always) testable of a large amount of data available (Anumba, 1995).

Regression models for prediction are often useful even when the assumptions are moderately violated, although they may not perform optimally. However, in many applications, especially with small effects or questions of casualty based on observational data, regression methods give misleading results.

Classical assumptions for regression analysis include:

i. The sample must be representative of the population for the prediction.

ii. The error is assumed to be a random variable with a mean of zero conditional on the explanatory variables.

iii. The variables are error-free. If this is not so, modeling may be done using errors-in-variables model techniques.

iv. The predictors must be linearly independent, that is, it must not be possible to express any predictor as a linear combination of the others.

v. The errors are uncorrelated, that is, the variance-covariance matrix of the errors is diagonal and each non-zero element is the variance of the error.

vi. The variance of the error is constant across observations (homoscedasticity). If not, weighted least squares or other methods might be used.

These are sufficient (but not all necessary) conditions for least-square estimator; in particular, these assumptions imply that the estimates would be unbiased, consistent and efficient in the class of linear unbiased estimators. Many of these assumptions may be relaxed in more advanced treatments (Cressie, 1996). Independent and dependent variables often refer to values measure at point locations. There may be spatial trends and spatial autocorrelation in the variables that violates statistical assumptions of regression.

Geographic weighted regression is one technique to deal with such data (Allen and Rober, 20002). Also, variables may include values aggregated by areas. With aggregated data, the Modifiable Area Unit Problem can cause extreme variation in regression parameters (Sridhar, 2003). When analyzing data aggregated by political boundaries, postal codes or census areas results may be different with a different choice of units.

MATERIALS AND METHODS

Data Used

The type of data collected for this paper is secondary data and is strictly based on the road traffic accident count in Nigeria for the space of ten (10) years (2001-2011) obtained from the Policy, Research and Statistics department of the Federal Road Safety Commission, Abuja, Nigeria (FRSC, 2012). After identification of a problem that is worth investigating, a research for the relevant sources and the design of method of collecting data and procedures became necessary.

The procedure involved in collecting road traffic accident data by the Federal Road Safety Commission as explained on interviewing the officer in the commission. The operations department of the commission is responsible for the collation of road traffic accident figures while the policy, research and statistics department takes the responsibility of collecting the figures together to make up the data. The Deputy Corps Marshal of this department heads the team.

The operations department carries out census on the highways, which is aimed at collecting information on the volume of accidents on the highways in the country. The staff of the department which are part of the census are distributed to various states of the federation where the census is meant to cover. The staff involved are instructed on how to carry out the census and each of them is given a sheet that contains all the categories of accidents that will be counted during the census for the sake of accuracy.

Relevant points to be noted in the census include; fatal cases, serious cases, minor cases, persons killed and persons injured. It is this information collected during the census from the various states of the federation that is put together to make up the 'road traffic accident summary' for each year as related by the Policy, Research and Statistical department.

Regression Models

Dale and Davies (1994) provided an excellent description of the sequence of procedures that articulate statistical modeling, namely, model formulation, model fitting, model criticism or assessment and model interpretation.

Regression models involve the following variables:

i. The unknown parameters denoted as β; may be a scalar or a vector of length k.

ii. The independent variable, X.

iii. The dependent variable, Y.

A regression model is a function that relates *Y* to X and β.

$$Y = f(x, \beta) \quad \textbf{(1)}$$

The approximation is usually formalized as:

$$E(Y/x) = f(x, \beta) \quad \textbf{(2)}$$

To carry out regression analysis, the form of the function f must be specified. Sometimes the form of this function is based on knowledge about the relationship between Y and X that does not rely on the data. If no such knowledge is available, a flexible or convenient form for *f* is chosen.

In the more general multiple regression model, there are *p* independent variables:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + e_i \quad \textbf{(3)}$$

where Y_i = Dependent variable.
X_i = Predictor, independent variable
β_i = Coefficient of regression
e_i = Residual errors

RESULTS AND DISCUSSION

The following are the results obtained after performing regression on the data used in this paper with the aid of Statistical Package for Social Sciences (SPSS).

In each of the results, the number of persons who died through accidents is taken as the dependent variable while the fatal cases, serious cases and minor cases are taken as the independent variables.

Table 1: Model Summary.

R	R Square	Adjusted R square	Std. Error of Estimate
0.767	0.589	-0.027	424.21568

Model obtained from result is;

$$Y = 615.929 + 1.735X_1 + 1.864X_2 - 0.797X_3 \quad \textbf{(4)}$$

Equation (4) can be used to predict the number of lives lost in road accident knowing the numbers of Fatal, Serious and Minor accidents on the road. From the model, it is evident that the intercept on the Y-axis is 615.929 while for every fatal road accident, there is an increase of about 173.5% of lives that would be lost. Similarly, for every one serious accident, there are about 186.4% increases in lives that would be lost and finally about 79.7% decreases in lives to be lost due to a minor road accident.

Table 2: ANOVA Table.

Source of Variation	Sum of Square (SS)	Degree of Freedom (df)	Mean Square (MS)	F	P
Regression	515883.439	3	171961.146	19.56	0.004
Error	59917.894	2	179958.947		
Total	875801.333	5			

Table 3: Coefficients of Regression.

Model	Unstandardized Coefficients		Standardized Coefficients	t	p
	B	std. Error			
Constant	615.296	19.418	1.714	0.137	0.000
Fatal Cases	1.121	0.138	0.747	8.129	0.000
Serious Cases	1.043	0.154	0.032	0.282	0.787
Minor Cases	-0.610	0.233	-0.274	-0.870	0.028

The model has a goodness of fit because the result satisfies all the necessary conditions for goodness of fit.

i. R-Square or the coefficient of determination. The value of R-Square shows that 58.9% of the variation was due to regression i.e. explained variation while the remaining 41.1% was unexplained or due to error.

ii. The P_value (0.004) that is less than α (0.05) is an indication of goodness of fit.

Therefore, the model is therefore valid and useful.

CONCLUSION

From the results obtained in this paper, it is obvious that the fatal kind of accident tends to be claiming the highest number of lives in majority of the states and in the entire country at large. This type of accident should therefore be given much attention and utmost priority in all the states.

RECOMMENDATIONS

In order to avoid the continuous loss of lives on our roads, this paper recommends that the following measures be put in place by the government;

i. Construction of lanes and highway shoulders: The Ministry of works should, to some extent, utilize the information on the traffic to increase the lanes on the road to allow free flow of traffic and over takings. Highway shoulders should also be constructed to aid motorists who want to stop and repair or check faults in their vehicles.

ii. Increasing the thickness of the road: Government should increase the thickness on some roads as the need arises to avoid total depletion of the road. Bad portions of the roads should also be patched. This is common to roads where trailers and trucks usually pass through frequently.

iii. Demarcation of roads: Road marking should be carried out. The lanes should be marked so as to assist the road users when driving not to go out of their lanes or drive between two lanes.

iv. Traffic control marshals should be assigned to some roads to control the system and clear accident scenes.

v. Also signs and signals, speed limits and warning instructions should be on the highways to regulate, enforce and co-ordinate all road traffic and safety management activities

REFERENCES

1. Ader, H.J. 2008. *Advising on Research Methods: A Consultant's Companion.*
2. Allen, H.K. and J.C. Robert. 2000. *Introduction to Business Statistics. 5th Edition.* South-Western College Publishing: Los Angeles, CA.
3. Anderson, D.R., D.J. Sweeney, and T.A. Williams. 1996. *Statistics for Business and Economics. 6th Edition*. South-Western College Publishing: Los Angeles, CA.
4. Andrew, W.E. 2003. "Estimating Transport Fatality Risk from Past Accident Data". *Journal of Science and Technology*. 2:7-14.
5. Anumba, J.U. 1995. *Economics and Time Series Analysis*. John Wiley and Sons: Montreal, Canada.
6. Daily Trust. 2010. www.dailytrustnewsonline.com
7. Dale, L.A. and L.I. Davies. 1994. *Statistical Method for Research Work*. Halted Press: New York, NY.
8. Devore, J. and R. Peck. 1986. *Statistics: The Exploration and Analysis of Data*. West Publishing Company: St. Paul: MN.
9. Federal Road Safety Commission. 2012. "FRSC Accidents Summary Statistics". Department of Policy, Research and Statistics. FRSC Publication Unit: Lagos, Nigeria.
10. Federal Road Safety Commission. 1988. "FRSC Act Cap 141 Laws of Federal Government of Nigeria". FRSC Publication Unit: Lagos, Nigeria.
11. Griffith, D.A. 2009. *Spatially Autoregressive models. 2nd Edition*. PWS-KENT Publishing Company: Boston, MA.
12. Gustavsson, J. 1969. "On the Use of Regression Models in the Study of Road Accidents". *International Journal of Science*.1:20-29.

13. Kapada, R. and G. Adersion. 1987. *Statistics Explained: Basic Concepts and Methods*. John Wiley and Sons: New York, NY.

14. Lindman, H.R. 1974. *Analysis of Variance in Complex Experimental Designs. 5th Edition*. Duxbury Press: Boston, MA.

15. Mason, R.D. 1992. *Statistical Techniques in Business and Economics. 10th Edition*. John Wiley and Sons: New York, NY.

16. Montgomery, D.C. 1994. *Applied Statistics and Probability for Engineers*. John Wiley and Sons: New York, NY.

17. Nicholson, A.J. 1985. "The Variability of Accident Counts". *Journal of Science*.1:37-42.

18. Nicholson, A.J. and Y.D. Wong. 1993. "Are Accidents Poisson Distributed?". *Journal of Environmental Technology*. 2:43-54.

19. Spiegel, M.R. and R.W. Boxes. 1972. *Theory and Problem of Statistics in S.I Units*. McGraw- Hill: London, UK.

20. Sridhar, M.S. 2003. *Data Collection Methods and Techniques*. Kin Keong Printing: Singapore.

SUGGESTED CITATION

Nafiu, L.A., D.I. Lanlege, and D. Hakimi. 2013. "Development and Implementation of an Empirical Model on Road traffic Accidents". *Pacific Journal of Science and Technology*. 14(2):170-175.