MIKAILALSYS
Journal of Multidisciplinary Sciences

e-ISSN : 2987-2286
p-ISSN : 2987-3924

Index : Dimensions, Scilit, Crossref, Garuda, Semantic, Lens,Google Scholar,Base, etc

https://doi.org/10.58578/mikailalsys.v3i2.5777

Application of Linear Probability Model to Road Traffic Crash

Adeniyi Oyewole Ogunmola[1], Victor Ajibo Ogebe[2], James Oruarooghene Onowuzou[3]
Federal University Wukari, Taraba State, Nigeria
ogunmola@fuwukari.edu.ng; victorogeber1@gmail.com

Article Info:

Submitted:	Revised:	Accepted:	Published:
Apr 6, 2025	Apr 27, 2025	May 12, 2025	May 18, 2025

Abstract

Road traffic crashes remain a critical public health and safety concern, particularly in developing countries such as Nigeria, where they constitute one of the leading causes of mortality and injury. This study investigates the likelihood that a road traffic crash in each of Nigeria's six geopolitical zones and in the country as a whole results in a minor incident. Quarterly data on road traffic crashes were sourced from the official database of the Federal Road Safety Corps and analyzed using a linear probability model. The model estimates the probability of a crash being categorized as minor across regions. Findings indicate that the probability of minor road traffic crashes is consistently below 20 percent in all zones and nationally, suggesting that the majority of reported crashes result in major damage or casualties. These results point to a concerning trend in crash severity across Nigeria. The study highlights the urgent need for enhanced traffic safety interventions, stricter enforcement of road regulations, improved vehicle and infrastructure standards, and more effective emergency response systems. Emphasizing preventative strategies and public awareness campaigns could help shift the balance toward more minor, less harmful outcomes when crashes do occur. Ultimately, the goal should be to ensure that in the event of a road traffic crash, the incident remains minor in nature, minimizing harm to life and property.

Keywords: Road traffic crash; Linear probability model; Minor crashes; Heteroscedasticity; Nigeria; Geopolitical zones.

INTRODUCTION

Each year, more than 39,000 Nigerians lose their lives in car accidents, according to the World Health Organization's 2018 report on global road safety, this report estimated that road traffic fatalities in Nigeria would reach 39,802. Despite Africa having the lowest motorization rate, it experiences the highest rate of road traffic accidents, as highlighted by WHO (2013a) and Sumaila (2013). With the rising death toll from road accidents, especially in developing nations, research has increasingly focused on safety studies over the last three decades.

The World Health Organization (2010) reports that almost 1.3 million individuals lose their lives annually due to road traffic collisions, equating to over 3,000 fatalities per day. Moreover, between twenty to fifty million people suffer non-fatal injuries from these collisions, contributing significantly to disabilities worldwide. Despite low- and middle-income countries having only about one-third of the world's registered vehicle fleet, they account for eighty percent of road traffic deaths, as highlighted in the WHO Report. Three main factors contribute to road traffic accidents (RTAs): human errors, vehicle-related issues, and road conditions (Asalor & Ovuwori, 2010). Additionally, weather conditions can play a crucial role in the severity of injuries sustained in road accidents (Onokala, 2015; Dong, Huang, Jiang, & Richards, 2015).

Adekunle (2010) suggests that understanding the direct costs of traffic accidents can be simplified by considering the labor lost to the country's economy and the subsequent decrease in productivity. Additionally, Enu (2014) emphasizes that road accidents have adverse effects on labor quality, leading to the depletion of human capital, contraction of the market, and hindrance to potential economic growth. These accidents result in annual economic losses equivalent to two to three percent of the GDP in low- and middle-income countries, totaling an estimated $518 billion globally.

The objective of this study is to assess whether the influence of geopolitical zones significantly affects the proportion of minor road traffic crashes in Nigeria using a linear probability model.

MATERIALS AND METHODS

Data Collection: Quarterly data was obtained from the official website of federal road safety corps Nigeria. The data was also collected on a quarterly and geo-political zone basis from second quarter of 2020 to fourth quarter 2022.

Linear probability model (LPM)

The Linear Probability Model (LPM) is a statistical model used in regression analysis for binary outcomes. It assumes that the probability of an event occurring (coded as 1) or not occurring (coded as 0) is a linear function of predictor variables.

Method of Analysis

The linear probability model for this study is given as

$$\rho_i = \beta_o + \beta_1 NW_i + \beta_2 NC_i + \beta_3 SE_i + \beta_4 SW_i + \beta_5 SS_i + \beta_6 Q_{2i} + \beta_7 Q_{3i} + \beta_8 Q_{4i}$$

$$+ \beta_9 Q_{5i} + \beta_{10} Q_{6i} + \beta_{11} Q_{7i} + \beta_{12} Q_{8i} + \beta_{13} Q_{9i} + \beta_{14} Q_{10i} + \beta_{15} Q_{11i} + \varepsilon_i$$

where;

ρ_i represents response proportion of the occurrence of minor RTC in the ith quarter,

β_o is the intercept term representing the predict probability when all the predictor variables NW, NC, SE, SW, SS, Q_2, Q_3, …, Q_{11} are all absent.

$\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are regression coefficients of North West, North Central, South East, South West and South South respectively, compared to North East, the reference category.

$\beta_6, \beta_7, \beta_8, \ldots, \beta_{15}$ are regression coefficient of the Second Quarter, Third Quarter, Fourth Quarter on to the Eleventh Quarter respectively, compared to the First Quarter, the reference category.

ε_i *is* the error term representing the unobserved factors affecting the outcome that are not accounted for by the model. Where each of the error terms is independently, identically and normally distributed with mean zero and constant variance.

Assumptions of Linear Probability Model (LPM)

The linear probability model makes several assumptions, including:

1. Linearity: Assumes a linear relationship between the explanatory variables and the probability of the event.

2. Independence: Assumes that observations are independent of each other.

3.Homoscedasticity: Assumes constant variance of the error term across all levels of the independent variables.

4.No Perfect Multicollinearity: Assumes that there is no perfect linear relationship between the independent variables.

5. Normality of Errors (optional): While not strictly necessary for estimation, the assumption of normally distributed errors can be useful for hypothesis testing

Diagnostic Checking

Diagnostic checking is used to evaluate the assumptions of a statistical model by assessing the fit of the model to the data, identifying new discrepancies or violation, and diagnosing potential issues to improve the models reliability and validity

Test for Normality

Testing for normality in statistics determines if a dataset conforms to a normal distribution, crucial for accurate parametric analysis. This assessment informs whether to use parametric or non-parametric methods, impacting the validity of statistical inferences. Identifying deviations from normality is essential for ensuring the reliability of statistical analyses and the robustness of research conclusions.

Test for Heteroscedasticity

Heteroscedasticity testing in statistics examines if the variance of errors in a regression model is consistent across predictor values. Detecting heteroscedasticity is vital as it affects the precision of parameter estimates and can lead to biased statistical inferences. Addressing heteroscedasticity through model adjustments or robust standard errors enhances the reliability of regression analyses and improves the validity of research findings.

RESULTS AND DISCUSSION

Descriptive Statistics

The descriptive statistics which include the mean, standard deviation and the number of observation (N) are presented in the Table 1 below:

Table 1: Descriptive Analysis of the Geo-Political Zones of Nigeria

Geo Political Zone	Mean	Std. Deviation	Sample Size
North East(NE)	357.18	60.183	11
North West(NW)	525.36	87.256	11
North Central(NC)	910.73	170.713	11
South East(SE)	166.82	26.339	11
South West(SW)	676.91	68.449	11
South South(SS)	170.64	24.109	11
Total	467.94	283.964	66

Table 1 above reveals the descriptive statistics of road traffic crashes (RTC) in the six geopolitical zones of the Nigeria. The result reveals north central having the highest mean (910.73) followed by south west (676.91), while south east as the lowest of (166.82)

Testing for the Validity of Model Assumptions

Normality Test for Residuals

Table 2: Tests of Normality of residuals of road traffic crashes (RTC)

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	Df	Sig.
Unstandardized Residual	.193	66	.000	.891	66	.000

Testing for the Validity of Model Assumptions of the Transformed Road Traffic Crashes

From table 1, the test for normality above reveal the residual road traffic crash not to be normal with (p=0.00) hence it had to undergo natural logarithm(Ln) transformation

Heteroskedasticity test

Testing for heteroscedasticity,

Tables 4: Modified Breusch-Pagan Test for Heteroskedasticity

Chi-Square	Df	Sig.
8.441	1	.004

Tables 5: Breusch-Pagan Test for Heteroskedasticity

Chi-Square	Df	Sig.
18.640	1	.000

From table 4 the result above reveal that the RTC data failed the heteroscedasticity test hence it needs to be transformed from Road traffic crashes to RTC natural logarithm.

Tables 6: Tests of Normality of Natural logarithm of road traffic crashes (RTC)

	Statistic	df	Sig.	Statistic	df	Sig.
Residual for Ln_RTC	.069	66	.200*	.990	66	.875

From Tables 6. The test for normality above reveal the Natural logarithm of road traffic crashes (RTC) to be normal with the significant value Kolmogorov-Smirnov to be 0.200 and Shapiro-Wilk as 0.875 hence it follows normality.

Test for heteroscedasticity

Table 7: Modified Breusch-Pagan Test for Heteroskedasticity

Chi-Square	Df	Sig.
.180	1	.671

Table 8: Breusch-Pagan Test for Heteroscedasticity[a b,c]

Chi-Square	df	Sig.
.209	1	.647

The result from Table 7 and Table 8 reveal that the Natural logarithm of road traffic crashes (RTC) data passed the heteroskedasticity test.

Table 9: Model Summary of Linear probability regression

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.927[a]	.859	.847	.0274731	1.929

Table 9 shows the model summary of the linear probability regression. It shows that the multiple correlation between the proportion of minor RTC and the set of geopolitical zones is a strong relationship with a value of 0.927, it shows that the R-squared and the adjusted R-squared values to be 0.859 and 0.847 respectively, the adjusted R-squared shows that about 84.7% of the variations in proportion of minor road traffic crash is explained by the predictors, also from 4.12: Durbin-Watson value is 1.929 which shows that there is no serial autocorrelations in the regression residuals that is their assumptions of independent in the residuals is not violated.

Table 10: Analysis of Variance for the Linear regression probability estimation

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.276	5	.055	73.208	.000b
	Residual	.045	60	.001		
	Total	.322	65			

a. Dependent Variable: Proportion of Minor RTC

Tables 10: shows the analysis of variance for the linear regression probability. The F-statistics value 73.208 and the corresponding p-value is 0.000, since the p-value is less than 0.05 level of significance, the null hypothesis of the linear regression model does not fit the data hence it rejected.

Table 11 shows the estimated parameters, their test statistics and p-values, and the collinearity statistics in the linear probability model. From the table, the collinearity tolerance is 0.6, constant for all the predictors. The Variance Inflation Factor (VIF) is at 1.667 is also constant for all the predictors. Since the tolerance is greater than 0.1, and the VIF is less than 10, there is no significant multi-collinearity in the predictors that must be corrected. Significant amount of information contained in one predictors is not contained in other predictors

Table 11: Coefficients of Linear Probability Regression

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	.021	.008		2.526	.014		
	North West	.004	.012	.023	.368	.714	.600	1.667
	North Central	.115	.012	.616	9.847	.000	.600	1.667
	South East	.138	.012	.736	11.769	.000	.600	1.667

South West	.164	.012	.874	13.974	.000	.600	1.667
South South	.126	.012	.674	10.781	.000	.600	1.667

a. Dependent Variable: Proportion of Minor RTC

$\hat{P}i$ =0.021+0.004NW+0.115NC+0.138SE+0.164SW+0.126SS

From the estimated equation above, where proportion of minor road traffic crashes is the dependent variable, 0.021 is the constant in the estimated in the probability linear model, when an accident occur in the North west (NW) zone the proportion of the accident being minor increases by 0.004, when an accident occur in the North Central (NC) zone proportion of the accident being minor increases by 0.115, when an accident occur in the South East (SE) zone the proportion of the accident being minor increases by 0.138, when an accident occur in the South west (SW) zone the proportion of the accident being minor increases by 0.164, when an accident occur in the South South (SS) zone the proportion of the accident being minor increases by 0.126.

Large Sample Approximations of 95% Confidence Interval for Proportions

Table 11 above shows the estimated proportion, standard error of estimated proportions and large sample approximations of 95% confidence interval proportion for each of the geo-political zones and the whole Nigeria, For North East zone, the estimated proportion of minor road traffic crash is 0.0209238 while the 95% confidence interval for proportion of a minor accident occurring in the zone lies between -0.0636 and 0.1054.

Table 12: Large Sample Approximations of 95% Confidence Interval for Proportions

Geo-Political Zone	N	Estimated Proportion	Standard error of Estimated Proportion	Lower Bound	Upper Bound
NE	11	0.0209238	0.043155	-0.0636	0.1054
NW	11	0.0252341	0.047287	-0.0675	0.1180
NC	11	0.1362751	0.103443	-0.0664	0.3390
SE	11	0.1587921	0.11019	-0.0572	0.3748
SW	11	0.1846203	0.11698	-0.0447	0.4139
SS	11	0.1472236	0.10683	0.0622	0.3567
Nigeria (Total)	66	0.1121782	0.09515	0.0360	0.1883

For North West zone, the estimated proportion of minor road traffic crash is 0.0252341 while the 95% confidence interval for proportion of a minor accident occurring in the zone lies between -0.0675, 0.1180. For North Central zone, the estimated proportion of minor road traffic crash is 0.1362751 while the 95% confidence interval for proportion of a minor accident occurring in the zone lies between -0.0664, 0.3390. For South East zone, the estimated proportion of minor road traffic crash is 0.1587921 the 95% confidence interval for proportion of a minor accident occurring in the zone lies between -0.0572, 0.3748. For South West zone, the estimated proportion of minor road traffic crash is 0.1846203 while the 95% confidence interval for proportion of a minor accident occurring in the zone lies between -0.0447, 0.4139. For South South zone, the estimated proportion of minor road traffic crash is 0.147236 while the 95% confidence interval for proportion of a minor accident occurring in the zone lies between 0.0622, 0.3567. For Nigeria as a country, the estimated proportion of minor road traffic crash is 0.1121782 while the 95% confidence interval for proportion of a minor accident occurring in the zone lies between 0.0360, 0.1883, The bound for north east, north west, north central, south east and south west zones are invalid since the contains negative value, hence the 95% confidence interval for the proportion using logit transformation is carried out to ensure valid bounds or intervals.

95% Confidence Interval for Proportions Using Logit Transformation

Table 13 below shows the 95% confidence interval for the proportion using logit transformation to ensure valid bounds or intervals.

Table 13: 95% Confidence Interval for Proportions Using Logit Transformation

Geo-Political Zone	Estimated Log-Odds	Lower Bounds of Estimated log-Odds	Upper Bounds of Estimated Log-Odds	Lower Probability Bound	Upper Probability Bound
NE	-1.6702	-5.8009	2.4569	0.0030	0.9248
NW	-1.5869	-5.3549	2.1811	0.0047	0.8985
NC	-0.8020	-2.5245	0.9205	0.0742	0.7151
SE	-0.7241	-2.3410	0.8929	0.0877	0.7095
SW	-0.6451	-2.1682	0.8780	0.1027	0.7064
SS	-0.76287	-2.4303	0.9050	0.0809	0.71197
Nigeria (Total)	-0.8983	-1.6582	-0.1294	0.1600	0.4677

Table 13: reveals the logit transformation for the 95% confidence interval for the proportion of a minor accident occurring. For North East zone, the 95% confidence interval for proportion of a minor RTC occurring in the zone lies between 0.0030 and 0.9211, For North West zone, the 95% confidence interval for proportion of a minor RTC occurring in the zone lies between 0.0047 and 0.8985. For North Central zone, the 95% confidence interval for proportion of a minor RTC occurring in the zone lies between 0.0742 and 0.7151. For South East zone, the 95% confidence interval for proportion of a minor RTC occurring in the zone lies between 0.0877 and 0.7095. For South West zone, the 95% confidence interval for proportion of a minor RTC occurring in the zone lies between 0.1027 and 0.7064.

Also for Nigeria as a whole, the 95% confidence interval for proportion of a minor RTC occurring in the zone lies between 0.0360 and 0.1883.

DISCUSSION

This project work investigated the probability of a minor accident occurring in each of the geopolitical zones and Nigeria as a whole. Linear probability model was employed to analyze the data, normality test and heteroscedasticity test were failed. Natural log transformation was applied and non-normality and heteroscedasticity were addressed.

Linear Probability Model (LPM) was employed to determine the effect of the geopolitical zones on the probability of minor accident occurring. Fitting LPM to the data, there is a strong correlation of about 0.927 between the estimated probability of minor RTC and the set of geopolitical zones. About 84.7% of the variations in proportion of minor road traffic crashes is explained by the geopolitical zones. Durbin-Watson statistic value was 1.929, which is very close to 2, that is, there is no serial autocorrelations in the regression residuals, the assumption of independence in the residuals is not violated. In the estimation of the probability of minor RTC, North East zone was taken as the based region. When a RTC occurs in North West zone, estimated probability of minor RTC increases insignificantly by 0.004 compared to North East zone. When a RTC occurs in North Central zone, estimated probability of minor RTC increases significantly by 0.115 compared to North East zone. When a RTC occurs in South East zone, estimated probability of minor RTC increases significantly by 0.138 compared to North East zone. When a RTC occurs in South west zone, estimated probability of minor RTC increases by

0.164 compared to North East zone. When a RTC occurs in South South zone, estimated probability of minor RTC increases significantly by 0.126 compared to North East zone. The estimated probability of minor RTC is highest for South West zone, followed by South East zone, followed by South South zone, followed by North Central zone. The lowest are those of North West zone and North East zone.

The estimated proportion of minor RTC occurred in the North East zone is about 2.0% of the total road traffic crashes in the zone, and its 95% confidence interval is (0.0030 and 0.9211). The estimated proportion of minor RTC occurred in the North West zone is about 2.5% of the total road traffic crashes in the zone, and its 95% confidence interval is (0.0047, 0.8985). The estimated proportion of minor RTC occurred in the South East zone is about 15.9% of the total road traffic crashes in the zone, and its 95% confidence interval is (0.0877, 0.7095). The estimated proportion of minor RTC occurred in the South West zone is about 18.5% of the total road traffic crashes in the zone, and its 95% confidence interval is (0.1027, 0.7064). The estimated proportion of minor RTC occurred in the South South zone is about 14.7% of the total road traffic crashes in the zone, and its 95% confidence interval is (0.0622, 0.3567). The estimated proportion of minor RTC occurred in Nigeria as a whole, is about 11.2% of the total road traffic crashes in the country, and its 95% confidence interval is (0.0360, 0.1883)

The range of the 95% confidence interval for the probability of minor road traffic crashes occurring in the North East is the largest followed by North West zone, North Central zone, South South, South East zone, and the least is South West. The narrower the range of the 95% confidence interval for the probability of minor road traffic crashes in a particular zone, the higher the estimated regression coefficient for that zone.

CONCLUSION

The results from the linear probability model indicate significant zonal disparities in minor road traffic crash occurrences across Nigeria, with the South West zone contributing the highest proportion and the North East and North West zones the lowest. The South East zone exceeds the South South in minor crash proportions, while the South South surpasses the North Central zone. Notably, the South West exhibits both the highest estimated regression coefficient and the narrowest 95% confidence interval, suggesting greater precision and influence in its contribution. Conversely, the North East zone has the widest confidence interval and the lowest coefficient, reflecting higher uncertainty and lower proportional impact. Nationally, minor road traffic crashes account for only 11.2% of all crashes, with no zone exceeding 20%—the South West being the highest at 18.5%. These findings highlight a crucial public safety gap and underscore the urgent need for policy interventions aimed at mitigating crash severity. Promoting conditions that increase the proportion of minor, rather than severe, crashes should become a priority in Nigeria's road safety strategies.

REFERENCES

Asalor, J. O., & Ovuwori, O. (2010). *Road Traffic Accident in Developing Countries* (Vol. 1). Joja Press.

Dong, C., Richards, S. H., Huang, B., & Jiang, X. (2015). Identifying the Factors Contributing to the Severity of Truck-Involved Crashes. *International Journal of Injury Control and Safety Promotion, 22*(2), 116–126. https://doi.org/10.1080/17457300.2013.879505

Enu, P. (2014). Road Traffic Accidents and Macroeconomic Conditions in Ghana. *Social and Basic Sciences Research Review, 2*(9), 374–393.

Hindle, G. A., & Hindle, T. (2011). Safety Cameras and Road Accidents: Effectiveness in Local Authority Areas in England. *Journal of the Operational Research Society, 62*, 1181–1188. https://doi.org/10.1057/jors.2010.97

Sumaila, A. (2013). Road Crashes Trends and Safety Management in Nigeria. *Journal of Geography and Regional Planning, 6*(3), 53–62. https://doi.org/10.5897/JGRP12.104

World Health Organization. (2010). *Global Plan for the Decade of Action for Road Safety 2011–2020*. World Health Organization.

World Health Organization. (2013). *Road Safety in the WHO African Region: The Facts 2013*. http://www.who.int/violence_injury_prevention/road_safety_status/2013/report/fact_sheet_afro.pdf

World Health Organization. (2018). *Global Status Report on Road Safety 2018*. https://apps.who.int/iris/handle/10665/276462