



Analysis of Monthly Road Accidents Using Instrumental Variable

Hamza Mansur¹ and Okafor U.L²

1. Department of Mathematics, Sule Lamido University Kafin Hausa, hamzamansir@gmail.com

2. Department of Mathematics, Nigerian Defence Academy Kaduna, linusokafor@gmail.com

Abstract

The issue of road accidents is an increasing problem in developing economy especially Nigeria. The number of deaths on our roads are increasing geometrically to higher percentage, thus regulating traffic movement on roads is an important task. The objective of the research is to make comparisons between two different forecasting techniques and suggest the better one. We used method of regression analysis to generate the results of the two different models (IV and RV). Among the two models IV has the highest coefficient of determination of 0.7959 and lowest standard error of estimate of 0.2041 compared to RV with 0.5447 and 0.4553 respectively, the results are compared using a graphical method proposed by Wilson (2015). He suggested plotting the individual fitted values against the actual observed values, if the point lies approximately along the line $y = x$, then the model is significant and recommended for accidents analysis.

Key words: Road Accidents, Instrumental Variable (IV), Reduced Variable (RV), Prone zone, Regression Analysis

1. Introduction

Traffic safety play key and integral role in a sustainable transportation development strategy, the main negative impact of modern transport system today is loss of life and injury as a result of road accidents. The success of traffic safety and highway improvement programs hinges on the analysis of accurate and reliable traffic accident, the probability of occurrence and severity can often be reduced by the application of proper accident data analysis and good roadway design features. The rapid increases in vehicle ownership in Nigeria during the past five years have placed considerable pressure on the road networks, their traffic and control devices, and on users of road facilities. Thus, there is need to develop mathematical models that will be used to analyze road accidents and make comparison to suggest the best model for road accident analysis. Many researchers have devoted their research to the area of road accidents, for example; Adekunle (2012) used regression method to analyzed road traffic accidents in Nigeria for the period 1975 – 2009, the statistical package used in the study was SPSS (Statistical Package for the social science). The model for the study is

$$\text{Log (CAS)}_t = a + b (YR_t - 1988) + c \log (CAS)_{t-1} \tag{1}$$

where $(CAS)_t$ represents casualties in year t , YR_t represents the year, $(CAS)_{t-1}$ represents casualties in year $t - 1$, a is the intercept while b and c are coefficients to be determined. the model shows that the two independent variables are statistically significant as shown by the $f = 3.289$ value which is less than the table value of 3.32 at 0.05 level of significance, this means that the model is only suitable for explanation and not prediction of road accidents rate. Murthy and Srinivasa (2015) used fractional factorial method for the sensitivity analysis of accident prediction model. The evaluation of sensitivity analysis indicated that average daily traffic (ADT), lane width (W), width of paved shoulder (PA), median(H) and their interactions (i.e., ADT-W, ADT-PA and ADT-H) has significant effects on numbers of accidents. Liyanage and Terrance (2015) they developed Poisson and Negative-Binomial models based on traffic police report data within the Galle police division for years 2011, 2012 and 2013. From the results of Negative Binomial model they found that, the key variables which cause the occurrence of accidents are experience of the driver (year of driver license issue), vehicle type, light condition and time of

the accident. Also, the decision tree results show that, road geometry with straight roads contributing to the highest number of accidents this due to the possibility that, when road is straight, drivers would tend to ride at high speeds. Gajendran *et.al* (2015), they discuss about three types of accident prediction model “System Dynamic Model, Fuzzy logic and Bayesian Method. The Complex, Dynamic and Non-linear interaction can be understood using system dynamic model. Fuzzy logic deals with occurrence of sets and elements, Bayesian refer to methods in probability and Statics which has held to model the interaction between road geometry, traffic characteristics and accident frequencies by means of linear regression model. Prithvi *et.al* (2013) they developed an accident prediction model that depict the effect of some of the major factors on accident causation on urban roads. The study showed that carriageway width, shoulder type, road condition, land use and composition of trucks in traffic stream were the most significant parameters affecting accidents on urban roads. Mixed type of non linear regression analysis was adopted and a best fit model was selected which was statistically and practically significant.

Lazim (2012) used regression method and developed road accident model to describe road accidents in Malaysia over the period of 1974 to 2007, the model shows that the variable of registered vehicle and population provide higher impact to the number of road accident. Alcides *et.al* (2020) used Poisson regression model to analyze the contributing factors of road fatal death in Angola. The analysis shows that the distribution of road fatal deaths and injuries in Angola varies according to various types of crash pattern. Atubi A. (2012) used time series and average models to analyze monthly road traffic accidents in selected Local Government of Lagos state, Nigeria. The research revealed that the most accidents prone Local Government areas are; Lagos Island, Lagos Mainland, Ikeja, Apapa, Eti – oja, and Ojo.

2 Materials and Methodologies

Accidents data covering a period of forty eight months were collected (2010-2015) from Federal Road Safety Corps (FRSC). The data were then summed up according to the particular month the accident occurred; the essence of this is to determine the effect of a particular month in the year. From the Traffic accidents data collected, some variables are selected for analysis and modeling. These are;

y is the number of accident

X₁ is the number of death

X₂ is the number of injured

X₃ is the number of passengers

X₄ is the number of people involved in the accident

The end result of a regression analysis study is often to generate a model that can be used to forecast or predict future values of the response variable given specified values of the predictor variables. The instrumental Variables method is a single equation method, being applied to one equation of the system at a time, it attained the reduction of the dependence of *u* and the explanatory variables, by using appropriate exogenous variables (as instruments) and it is outline as follows.

From the structural form of, $y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + u$ (2)

2.1 Equations

We may multiply all the variables of the structural equation by each of the instrumental variables (Z_1, Z_2, Z_3, Z_4) and sum over all sample observations to obtain

$$\sum Z_1 y = a_1 \sum Z_1 X_1 + a_2 \sum Z_1 X_2 + a_3 \sum Z_1 X_3 + a_4 \sum Z_1 X_4 + \sum Z_1 u \quad (3)$$

$$\sum Z_2 y = a_1 \sum Z_2 X_1 + a_2 \sum Z_2 X_2 + a_3 \sum Z_2 X_3 + a_4 \sum Z_2 X_4 + \sum Z_2 u \quad (4)$$

$$\sum Z_3 y = a_1 \sum Z_3 X_1 + a_2 \sum Z_3 X_2 + a_3 \sum Z_3 X_3 + a_4 \sum Z_3 X_4 + \sum Z_3 u \quad (5)$$

$$\sum Z_4 y = a_1 \sum Z_4 X_1 + a_2 \sum Z_4 X_2 + a_3 \sum Z_4 X_3 + a_4 \sum Z_4 X_4 + \sum Z_4 u \quad (6)$$

The last term can be substituted by their zero expected value and hence dropped from the equations. If one of the explanatory variables, say X_1 , is exogenous we use it as an instrument for itself, that is set $Z_1 = X_1$, and follow the same procedure. This method involved the solution of transformed normal equations and each normal equation is multiplied through by an instrumental variable as;



$$\sum Z_1 y = \hat{a}_1 \sum Z_1 X_1 + \hat{a}_2 \sum Z_1 X_2 + \hat{a}_3 \sum Z_1 X_3 + \hat{a}_4 \sum Z_1 X_4 \tag{7}$$

$$\sum Z_2 y = \hat{a}_1 \sum Z_2 X_1 + \hat{a}_2 \sum Z_2 X_2 + \hat{a}_3 \sum Z_2 X_3 + \hat{a}_4 \sum Z_2 X_4 \tag{8}$$

$$\sum Z_3 y = \hat{a}_1 \sum Z_3 X_1 + \hat{a}_2 \sum Z_3 X_2 + \hat{a}_3 \sum Z_3 X_3 + \hat{a}_4 \sum Z_3 X_4 \tag{9}$$

$$\sum Z_4 y = \hat{a}_1 \sum Z_4 X_1 + \hat{a}_2 \sum Z_4 X_2 + \hat{a}_3 \sum Z_4 X_3 + \hat{a}_4 \sum Z_4 X_4 \tag{10}$$

The estimates of the structural parameters have the followings properties.

- for small samples the estimates are biased
- for large samples the estimates are consistent
- The instrumental variables estimates, although consistent are not asymptotically efficient; that is, they do not possess the minimum variance as compared with other consistent estimates obtained from alternative techniques.

The assumptions of method of instrumental variable can be outline as follows:

Firstly, knowledge of some exogenous variables in other equations of the complete system, which may be used as instruments.

Secondly, the new random term Zu must satisfy the usual assumptions of Ordinary Least Squares (OLS).

Thirdly, the exogenous variables (instrumental variables) must not be perfectly multi – collinear.

Fourthly, the structural function must be identified.

Fifthly, the variables must be properly aggregated.

2.2 Tables and Figures

We used Method of Instrumental Variable and reduced method analysis to test the relationship between accident Parameters and rate of accident. Also, we employed NCSS software that takes data from excel file and generates tabulated reports, Chart and estimated model. For each model, regression coefficients, regression equation, analysis of variance, normality test, Predicted and actual values were found.

Table 1: Residual Report for IV Model

Row	Actual Y	Predicted \hat{Y}	e_t Residual	$(e_t - e_{t-1})^2$	Absolute Percent Error	Sqrt (MSE) Without This Row
1	7.000	7.534	-0.534	0.285	7.626	1.512
2	0.000	1.194	-1.194	0.436	-	1.502
3	9.000	8.454	0.546	3.028	6.065	1.511
4	4.000	5.172	-1.172	2.952	29.298	1.503
5	4.000	3.532	0.468	2.690	11.697	1.512
6	3.000	4.003	-1.003	2.164	33.429	1.506
7	4.000	4.708	-0.708	0.087	17.705	1.510
8	5.000	4.366	0.634	1.801	12.677	1.511
9	0.000	1.194	-1.194	3.342	-	1.502
10	4.000	3.631	0.369	2.443	9.236	1.513
11	5.000	4.768	0.232	0.019	4.639	1.514
12	9.000	6.409	2.591	5.565	28.785	1.458
13	7.000	6.100	0.900	2.859	12.852	1.507
14	7.000	6.512	0.488	0.170	6.978	1.512
15	8.000	6.024	1.976	2.214	24.698	1.478



16	7.000	6.807	0.193	3.179	2.754	1.514
17	7.000	6.069	0.931	0.545	13.303	1.507
18	7.000	5.752	1.248	0.098	17.833	1.501
19	8.000	8.217	-0.217	2.146	2.712	1.514
20	9.000	5.839	3.161	11.411	35.121	1.429
21	8.000	8.420	-0.420	12.824	5.256	1.512
22	0.000	1.194	-1.194	0.599	-	1.502
23	7.000	4.676	2.324	12.376	33.199	1.468
24	15.000	13.016	1.984	0.116	13.228	1.465
25	0.000	1.194	-1.194	10.100	-	1.502
26	7.000	5.217	1.783	8.863	25.471	1.487
27	7.000	7.133	-0.133	3.671	1.896	1.514
28	8.000	10.387	-2.387	5.081	29.841	1.462
29	4.000	4.789	-0.789	2.554	19.733	1.509
30	6.000	5.082	0.918	2.914	15.300	1.507
31	6.000	6.033	-0.033	0.904	0.555	1.514
32	7.000	5.838	1.162	1.428	16.598	1.503
33	8.000	9.045	-1.045	4.871	13.067	1.504
34	7.000	7.810	-0.810	0.055	11.574	1.508
35	10.000	8.634	1.366	4.735	13.655	1.496
36	7.000	12.601	-5.601	48.539	80.014	1.116
37	8.000	8.078	-0.078	30.539	0.974	1.514
38	3.000	2.840	0.160	0.057	5.346	1.514
39	3.000	2.925	0.075	0.007	2.507	1.514
40	4.000	3.670	0.330	0.065	8.240	1.513
41	3.000	3.096	-0.096	0.181	3.197	1.514
42	7.000	6.310	0.690	0.618	9.852	1.510
43	5.000	5.995	-0.995	2.839	19.901	1.503
44	6.000	6.083	-0.083	0.832	1.377	1.514
45	0.000	1.194	-1.194	1.234	-	1.502
46	0.000	1.194	-1.194	0	-	1.502
47	8.000	8.063	-0.063	1.279	0.789	1.514
48	0.000	1.194	-1.194	1.279	-	1.502

$$\sum(e_t)^2 = 96.267 \quad \sum(e_t - e_{t-1})^2 = 205.994$$

The residuals (errors) of the model in table 1 were got by subtracting the predicted (\hat{Y}) from the actual (Y), where the absolute percentage error is the total residual divide by n multiplied by 100.

Table 2: Residuals report for reduced Model

Row	Actual Y	Predicted \hat{Y}	Residual e_t	Standard Error of Predicted	$(e_t - e_{t-1})^2$	95% Lower Pred. Limit of Individual	95% Upper Pred. Limit of Individual
2	0.000						
3	9.000	6.047	2.953	1.487	8.720	3.047	9.047
4	4.000	4.948	-0.948	1.460	15.218	2.002	7.894
5	4.000	5.268	-1.062	1.438	0.013	2.367	8.170
6	3.000	5.062	-2.062	1.430	1	2.177	7.948
7	4.000	5.284	-1.284	1.432	0.605	2.395	8.173



8	5.000	5.426	-0.426	1.421	0.736	2.558	8.294
9	0.000	4.018	-4.018	1.424	12.902	1.144	6.891
10	4.000	5.411	-1.411	1.461	6.796	2.463	8.359
11	5.000	5.435	-0.435	1.411	0.953	2.587	8.282
12	9.000	6.799	2.201	1.408	6.948	3.957	9.641
13	7.000	5.869	1.131	1.423	1.145	2.997	8.741
14	7.000	6.070	0.930	1.404	0.040	3.237	8.903
15	8.000	6.580	1.420	1.404	0.240	3.746	9.413
16	7.000	6.126	0.874	1.406	0.298	3.287	8.964
17	7.000	6.258	0.742	1.400	0.017	3.433	9.082
18	7.000	6.320	0.680	1.399	0.004	3.497	9.143
19	8.000	6.949	1.051	1.400	0.138	4.124	9.774
20	9.000	7.568	1.432	1.408	0.145	4.727	10.410
21	8.000	6.995	1.005	1.410	0.182	4.150	9.840
22	0.000	1.909	-1.909	1.425	8.491	-0.967	4.784
23	7.000	7.119	-0.119	1.449	3.204	4.194	10.043
24	15.000	12.420	2.580	1.491	7.285	9.412	15.429
25	0.000	0.984	-0.984	1.541	12.702	-2.126	4.095
26	7.000	7.306	-0.306	1.475	0.460	4.328	10.284
27	7.000	6.883	0.117	1.405	0.179	4.048	9.718
28	8.000	7.780	0.220	1.403	0.011	4.948	10.612
29	4.000	4.344	-0.344	1.402	0.318	1.515	7.174
30	6.000	6.384	-0.384	1.401	0.002	3.556	9.212
31	6.000	6.278	-0.278	1.396	0.011	3.460	9.096
32	7.000	7.265	-0.265	1.401	0.002	4.437	10.093
33	8.000	8.242	-0.242	1.414	0.005	5.389	11.095
34	7.000	7.251	-0.251	1.410	0.008	4.405	10.097
35	10.000	10.514	-0.514	1.448	0.069	7.591	13.437
36	7.000	7.237	-0.237	1.431	0.077	4.349	10.126
37	8.000	8.611	-0.611	1.422	0.140	5.742	11.480
38	3.000	2.969	0.031	1.423	0.412	0.097	5.840
39	3.000	3.259	-0.259	1.431	0.084	0.372	6.146
40	4.000	4.395	-0.395	1.418	0.018	1.533	7.257
41	3.000	3.076	-0.076	1.423	0.102	0.205	5.947
42	7.000	8.098	-1.098	1.440	1.044	5.193	11.004
43	5.000	5.319	-0.319	1.422	0.607	2.449	8.189
44	6.000	6.772	-0.772	1.426	0.205	3.894	9.651
45	0.000	-1.315	1.315	1.502	4.356	-4.347	1.717
46	0.000	-1.045	1.045	1.538	0.073	-4.149	2.060
47	8.000	10.021	-2.021	1.514	9.400	6.965	13.076
48	0.000	-1.893	1.893	1.540	15.319	-5.000	1.214

$$\sum(e_t)^2 = 75.417 \quad \sum(e_t - e_{t-1})^2 = 120.6768$$

Plot Section for Method of Instrumental Variable

$$C_2 = Y, C_3 = X_1, C_4 = X_2, C_5 = X_3, C_6 = X_4$$



Figure 1(a)

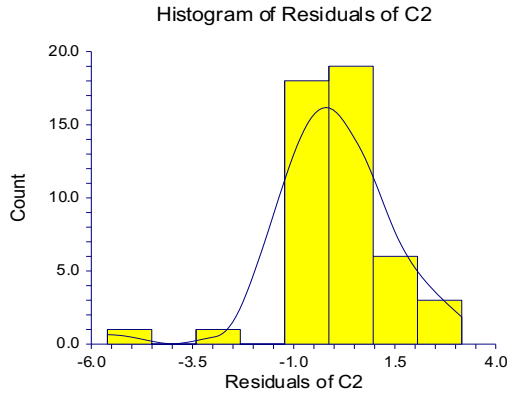


Figure 1(b)

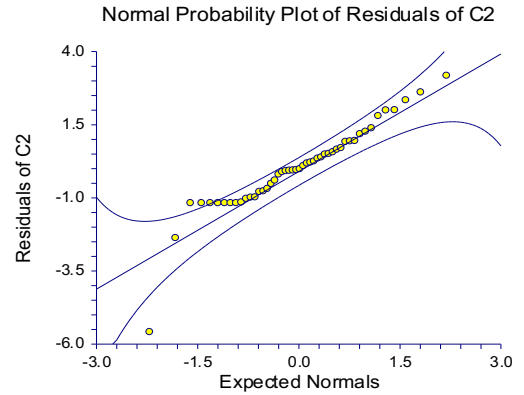


Fig 1(c)

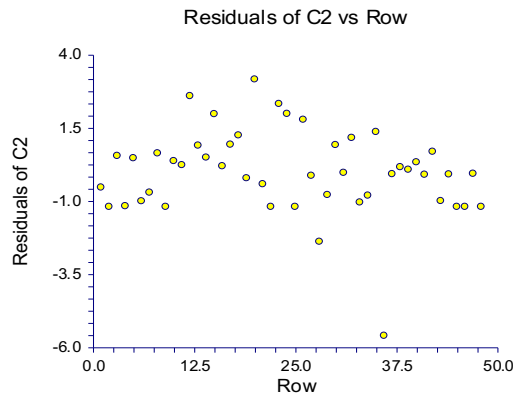


Fig 1(d)

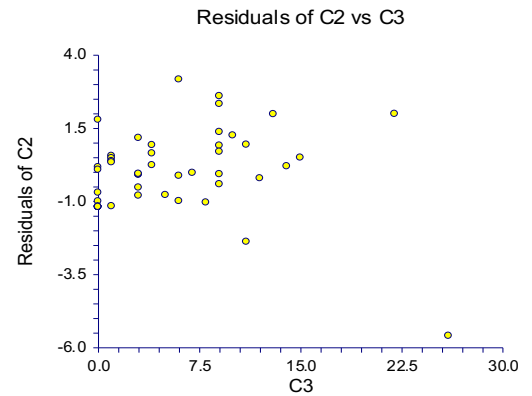


Figure 1(e)

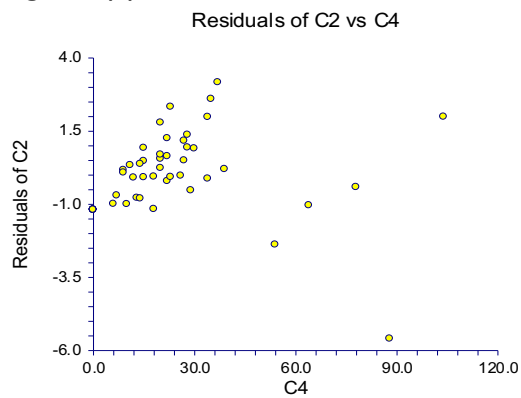


Figure 1(f)

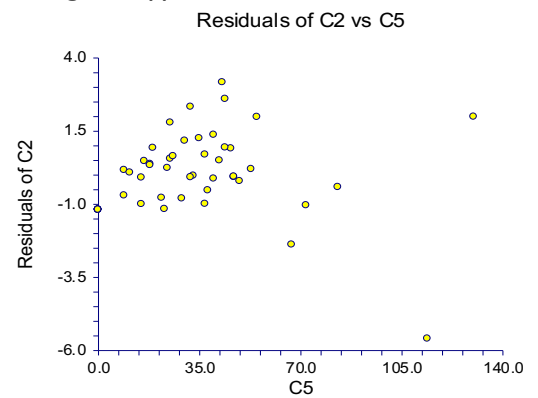
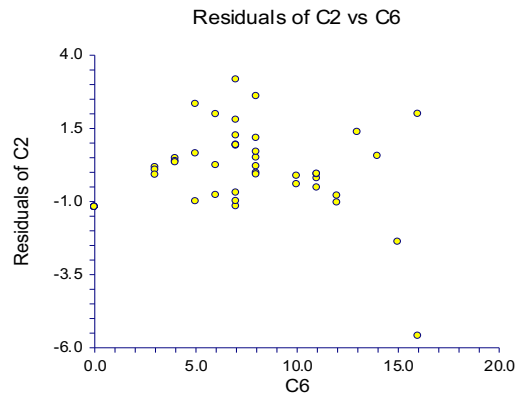


Figure 1(g)



Plot Section for Reduced Variable Model

$$C_1 = Y, C_2 = X_1, C_3 = X_3, C_4 = X_4$$

Figure 2(a)

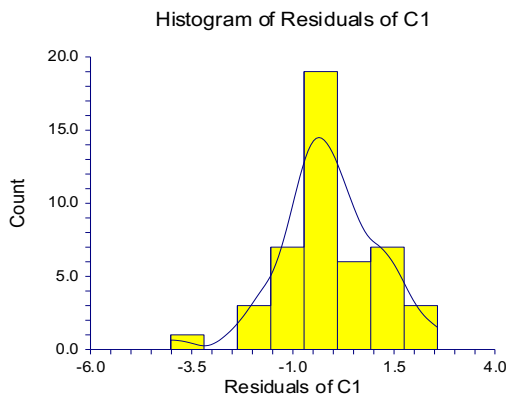


Figure 2(b)

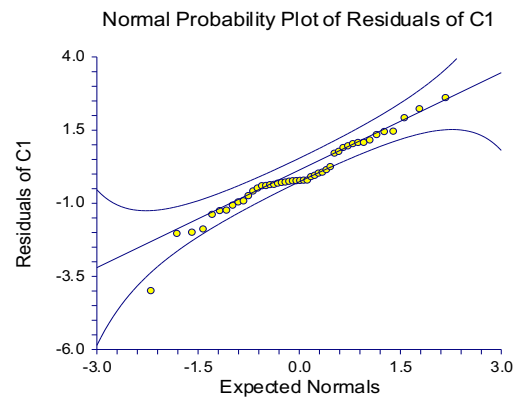


Figure 2(c)

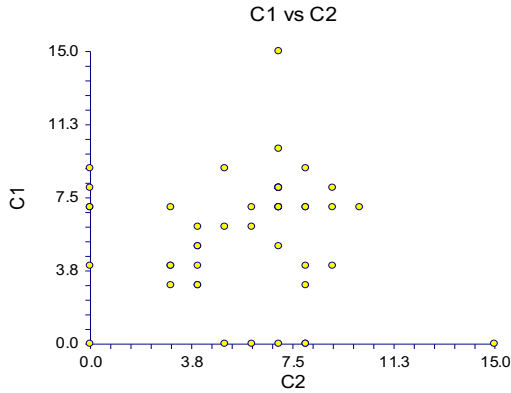


Figure 2(d)

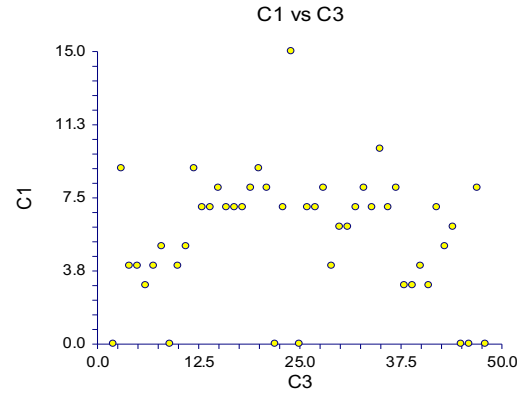


Figure 2(e)

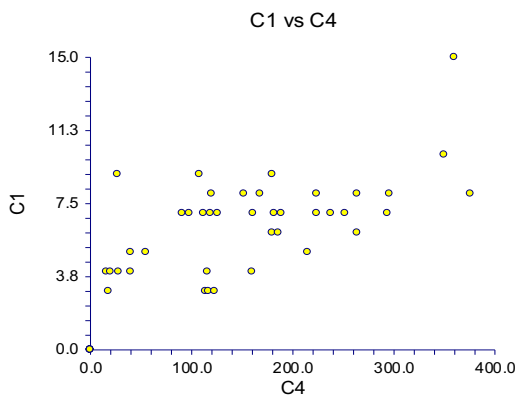


Figure 2(f)

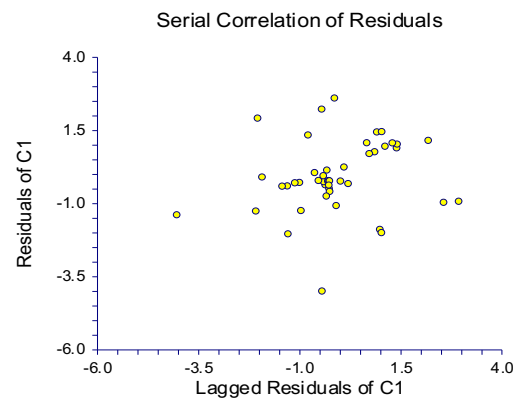


Figure 2(a) is a histogram of the residuals versus the residuals of y while figure 2(b) is the normal probability plot where all the variables are in expected normal. Figure 2(c), 2(d), 2(e) and 2(f) are the plot of residuals for detecting serial correlation where all the plots show the sign of serial correlation among the variables.

Run Summary Section Parameter	Value	Parameter	Value
Dependent Variable	Y	Rows Processed	48
Number Ind. Variables	4	Rows Filtered Out	0
Weight Variable	None	Rows with X's Missing	0
R2	0.7959	Rows with Weight Missing	0
Adj R2	0.7769	Rows with Y Missing	0
Coefficient of Variation	0.2680	Rows Used in Estimation	48
Mean Square Error	2.238749	Sum of Weights	48.000
Square Root of MSE	1.496245	Completion Status	Normal Completion
Ave Abs Pct Error	27.270		

3. Results

Method of Instrumental Variable (IV)

From the above table the R^2 of 0.7959 showed that the model is able to account for 79.59% of the change in the accidents (y) with only 20.41% not being explained by the model.



Descriptive Statistics Section

Variable	Count	Mean	Standard Deviation	Minimum	Maximum
X_1	48	5.625	5.902127	0	26
X_2	48	23.75	22.14892	0	104
X_3	48	33.20833	27.20134	0	130
X_4	48	6.979167	4.334373	0	16
Y	48	5.583333	3.16788	0	15

The descriptive statistic section with mean and standard deviation of each variable with X_3 has the highest mean of 33.20833 and standard deviation of 27.20134 which highly differ with dependent variable y that has the mean of 5.583333 and standard deviation of 3.16788 which means that X_3 has low effect in the change of y.

Regression Equation Section

Regression Independent Variable	Standard Coefficient b(i)	Error Sb(i)	T-Value to test $H_0: B(i)=0$	Prob. Level	Reject H_0 at 5%?	Power of Test at 5%
Intercept	1.1943	0.4149	2.879	0.0062	Yes	0.8034
X_1	0.0135	0.0859	0.158	0.8755	No	0.0527
X_2	-0.0133	0.0373	-0.355	0.7240	No	0.0640
X_3	0.0426	0.0446	0.955	0.3449	No	0.1543
X_4	0.4605	0.0930	4.954	0.0000	Yes	0.9980

The Estimated Model of Instrumental variable method

$$\hat{Y} = 1.1943 + 0.0135X_1 - 0.0133X_2 + 0.0426X_3 + 0.4605X_4 \tag{11}$$

The t – value of regression equation section above is used to test the significant of each independent variable in the model. If the t – value of the independent variable is less than the table value used to calculate the confidence limit then the variable is not significant in the model. And is calculate as $\frac{b(i)}{sb(i)}$

Regression Coefficient Section showing Lower and Upper 95% Confidence Limits

Independent Variable	Regression Coefficient	Standard Error	Lower 95% C.L.	Upper 95% C.L.	Standardized Coefficient
Intercept	1.1943	0.4149	0.3576	2.0309	0.0000
X_1	0.0135	0.0859	-0.1597	0.1867	0.0252
X_2	-0.0133	0.0373	-0.0886	0.0620	-0.0928
X_3	0.0426	0.0446	-0.0473	0.1325	0.3657
X_4	0.4605	0.0930	0.2730	0.6480	0.6301

Note: The T-Value used to calculate these confidence limits was 2.017

From the above table the lower and upper confidence limit at 95% is found from the mean of the dependent variable (y) at the value of each independent variable for the value of alpha.

Analysis of Variance Section

Source	DF	R2	Sum of Squares	Mean Square	F-Ratio	Prob. Level	Power (5%)
Intercept	1	-	1496.333	1496.333	-	-	-
Model	4	0.7959	375.4005	93.85011	41.921	0.0000	1.0000
Error	43	0.2041	96.26622	2.238749	-	-	-
Total (Adjusted)	47	1.0000	471.6667	10.03546	-	-	-

The Analysis of variance section of the model was described with sum of squares and mean square. The mean square is sum of square divided by the degree of freedom, while the f – ratio is the mean square total divided by mean square error.

Normality Tests Section

Test Name	Test Value	Prob. Level	Reject H0 At Alpha = 20%?
Shapiro Wilk	0.9185	0.002630	Yes
Anderson Darling	0.7921	0.040097	Yes
D'Agostino Skewness	-2.4344	0.014917	Yes
D'Agostino Kurtosis	3.1315	0.001739	Yes
D'Agostino Omnibus	15.7326	0.000383	Yes

From the normality test section, the entire test rejects the null hypothesis (that is the slope is zero) for normality test at alpha = 20%, means the errors are normally distributed.

Reduced Variable Method

Excluding the irrelevant Variable (X_2) in the Analysis we also Regressed the remaining three Variables using reduced variable Method and the results are shown in the tables below.

Run Summary Section

Parameter	Value	Parameter	Value
Dependent Variable	Y	Rows Processed	48
Number Ind. Variables	3	Rows Filtered Out	0
Weight Variable	None	Rows with X's Missing	1
R2	0.8251	Rows with Weight Missing	0
Adj R2	0.8126	Rows with Y Missing	0
Coefficient of Variation	0.3004	Rows Used in Estimation	47
Mean Square Error	1.896333	Sum of Weights	46.000



Square Root of MSE	1.377074	Completion Status	Normal Completion
Ave Abs Pct Error	24.865	Autocorrelation (Rho)	0.1921

When we remove the irrelevant variables in the model, the R^2 increases from 0.7959 to 0.8251 in the new formulated model, this means that including the irrelevant variable in the model make R^2 to decrease which make prediction to be less accurate.

Regression Equation Section

Independent Variable	Regression Coefficient b(i)	Standard Error Sb(i)	T-Value to test H0:B(i)=0	Prob. Level	Reject H0 at 5%?
Intercept	5.6805	0.6752	8.414	0.0000	Yes
X_1	-0.0694	0.0654	-1.061	0.2949	No
X_3	-0.1462	0.0205	-7.118	0.0000	Yes
X_4	0.0298	0.0022	13.568	0.0000	Yes

The Estimated Model

$$\hat{Y} = 5.6805 - 0.0694X_1 - 0.1462X_3 + 0.0298X_4 \quad (12)$$

Still X_4 is the most significant variable in the model as its t – value of 13.568 is greater than the table value of 2.018 used to calculate confidence limit and also there is variable that no need to be included in the model as the probability level of X_1 used to accept the null hypothesis.

Serial Correlation of Residuals from Uncorrected Model

Lag	Serial Correlation	Lag	Serial Correlation	Lag	Serial Correlation
1	-0.0639	9	0.0290	17	-0.0574
2	0.0840	10	-0.1899	18	-0.1209
3	0.1275	11	-0.1212	19	-0.1812
4	0.1375	12	-0.0097	20	0.0215
5	0.1025	13	-0.0788	21	0.0659
6	-0.2021	14	-0.1149	22	-0.1351
7	0.1251	15	-0.1408	23	-0.1041
8	-0.0051	16	-0.0083	24	0.0003

The Above serial correlations is significant if their absolute values are greater than 0.291730

Serial Correlation of Residuals from Corrected Model

Lag	Serial Correlation	Lag	Serial Correlation	Lag	Serial Correlation
1	-0.0222	9	0.0329	17	0.0409
2	-0.0121	10	-0.1174	18	-0.0245
3	0.0787	11	-0.0959	19	-0.1514
4	0.0495	12	0.0670	20	-0.0468
5	0.0782	13	0.0030	21	0.1790
6	-0.3212	14	-0.0708	22	-0.0136
7	-0.0662	15	-0.1286	23	-0.1802
8	-0.0903	16	0.0768	24	0.0583

For the two models there is no serial correlation of the residuals as their absolute values for serial correlation are less than the value for significant correlation (0.291730, 0.294884) respectively.

The 95% lower confidence interval is estimated for the mean of Y at the value of X, while the upper limit is also estimated for the mean of Y at each value of X after choosing the alpha level on the variables panel. The $\sum(e_t)^2$ and $\sum(e_t - e_{t-1})^2$ are used to find the value of d for Durbin Watson test of auto – correlation.

4 Discussion

It has been recognized that the most effective means towards accidents reduction lies on a systematic and scientific approach based on the use of accurate and reliable traffic accident data. From model eqn (11) the negative sign in the coefficient of X_2 indicates that the variable (X_2) has low correlation with dependent variable y (number of accident) and in regression equation section of the IV model the t – value of X_4 is statistically significant as its value of 4.954 is greater than the table value of 2.017 used to calculate the confidence limit. However, from the results of the two methods (IV and RV) X_4 is statistically significant which indicates that, the increase in the coefficients of X_4 (number of vehicles) will cause the rate of accidents to be higher. Even though, the reduced variable method has the highest value of 0.8251 for the R^2 , the model is inferred to the method of Instrumental variable because it produce auto – correlation errors. The analysis of variance section shows the performance of IV model in terms of the various tests for auto – correlation. Also, the study reveals that, a large value of R^2 does not necessarily mean high predictability no does low R^2 necessarily mean poor predictability because the IV model with low R^2 happen to be more consistent than the RV model with highest R^2 . It can be seen that adding an extra variable to the existing variable in the reduced model caused R^2 to decrease from 0.8251 to 0.7959 in the IV model showing that the additional variable does not contribute any significant to the dependent variable.

Contrary to claiming in literature by Adekunle (2012) where he used regression method $(\text{Log } (CAS)t = a + b(YR_t - 1988) + c \log (CAS)t - 1)$ to analyzed road traffic accidents in Nigeria using Statistical Package for the Social Science (SPSS), the model shows that the two independent variables are statistically significant as shown by the f = 3.289 value which is less than the table value of 3.32 at 0.05 level of significance, this means that the model is only suitable for explanation and not prediction of road accidents rate. In this paper two models were developed (IV and RV) and the better one (IV) was choose base on the results for prediction.

5 Conclusions

In this paper, method of regression analysis was used to develop two models (eqn.11 and eqn. 12) for analyzing the rate of road accidents occurrence. From the results of the two models the IV model with low R^2 is more consistent than the RV model with higher R^2 because it produced auto – correlation errors and perform well with less error of prediction from the various test conducted. Also, we find out that the increase in number of vehicles contributed

significantly to the increase in the rate of accidents and specifically the increase in number of passengers it do not contribute significantly to the change in the number of accidents. The results were compared with help of graphs and the probability plot of the residuals indicates that the individual probabilities of the residuals are normally distributed since the points are near or closer to one another with the plotted points forming an approximately straight line. However, the two models were developed using linear equation the work could be expand using non – linear equation and can be used by road safety authorities.

Acknowledgement

I wish to acknowledge the support, guidance and advice given by Wing commander UL Okafor (rtd), Ph.D Department of Mathematics, Nigerian Defence Academy Kaduna who contributed toward making this paper attain its present level. Also, I like to appreciate the support and encouragement rendered by Federal Road Safety Commission Jigawa state Headquarter, Dutse.

References

- Adekunle J. Aderamo (2012), assessing the trends in road traffic accidents casualties. *Journal of social science* 31(3) 19 – 25
- Alcides paulo, Baptista Lucas, Kananay Linjun (2020), Crash analysis and modeling of National wide road traffic Networks in Angola. *International Journal of traffic and transportation Engineering*, p –ISSN: 2325 – 0062, e – ISSN: 2325 – 0070 October 2020 9(1): 16 – 23.
- Atubi Augustus o. (2012), A monthly Analysis of road traffic accidents in selected local Government areas of Lagos State. *Mediterranean Journal of social Science*, Vol 3(11) November 2012.
- Gajendran C, Serin VK, Seenu SG, Swati P (2015), Different methods of accident forecast Base on real data. *Journal of civil environmental engineering* 5(4) 43 - 49
- Lazim, A. (2012), fuzzy linear regression model in describing road accidents in Malaysia over the period of 1974 to 2007; *International Journal of Scientific and Research Publications*, Volume 5(1): 250-253
- Liyanage T, and Terrance M (2015), Traffic accident analysis and development of accident prediction model. *Annual sessions of IESL* pp 105 - 109
- Murthy N. and Srinivavasa J. (2015), Development of model for road accidents based on ntersection parameters using regression models; *international Journal of scientific and research publications*, Volume 5, Issue 1, January 2015 ISSN 2250 – 3153
- Prithvi B, Lokesh H, Anantha R, priyanka K, (2013), Accident prediction modeling for an urban road of Bangalore. *International journal of Research in Engineering and Technology* ISSN: 2319 - 1163 .