NIPES

Advances in Engineering Design Technology

Journal homepage: www.nipesjournals.org.ng

Assessment of the Statistical Properties of Road Accident Data

Iyeke, S.D; Ilaboya, I.R and Ehiagbonare, M. O

Department of Civil Engineering, Faculty of Engineering, University of Benin, P.M.B 1154, Benin City, Edo State, Nigeria

Email of Corresponding Author: rudolph.ilaboya@uniben.edu

ARTICLE INFORMATION

Article history:
Received 21 December 2021
Revised 28 February 2022
Accepted 07 March 2022
Available online 27 March 2022

Keywords:
Cronbach alpha coefficient (CAC), test of normality, homogeneity test, and diagnostic statistics

https://doi.org/10.37933/nipes.a/4.1.2022.5

https://nipesjournals.org.ng

ABSTRACT

The soaring number of vehicles on the road had created a major social problem through traffic accidents due to the loss of lives and materials. Road traffic injuries cause considerable economic losses to individuals, their families, and to nations as a whole. The target of the study was to conduct a comprehensive statistical analysis of primary road accident data in order to assess their suitability for use in road accident prediction models. The study area is Ugbowo-Lagos Road in Benin City, Edo State Nigeria. A reconnaissance survey was done first to ascertain the geometric characteristic of the road which include; the chainage, the vertical and horizontal curve and the super elevation. Thereafter, primary data which include road accident data was collected from Federal Road Safety Office in Benin City and selected statistical techniques, namely; outlier detection, test of normality and autocorrelation test were conducted to assess the qualities of the data. With a computed p-value greater than 0.05 for all the independent variables, the null hypothesis of the Dixon test was accepted and it was concluded that the accident data obtained from FRSC is devoid of outliers. In addition, with a centered VIF < 10, it was concluded that there is the absence of multicollinearity between the dependent and independent variables. In addition, the calculated Cronbach alpha value of 0.900 as observed in the reliability test revealed that the data are reliable and the computed goodness of fit statistics of reliability gave a maximum Guttman coefficient of 88.10%.

1. Introduction

Every year, the lives of approximately 1.35 million people are cut short as a result of a road traffic crash [1]. Between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability as a result of their injury [2]. Road traffic injuries cause considerable economic losses to individuals, their families, and to nations as a whole. These losses arise from the cost of treatment as well as loss of productivity for those killed or disabled by their injuries, and for family members who need to take time off work or school to care for the injured. Road traffic crashes cost most countries 3% of their gross domestic product [3, 4 and 5].

The main purpose of transportation system is to provide the efficient and safe movement of freight and passenger from one place to another. The economic development is directly and strongly related to the availability of transportation. The soaring number of vehicles on the road had created a major social problem through traffic accidents due to the loss of lives and material [6]. Statistical or crash prediction model have frequently been used in highway safety studies. They can be used to identify

major contributing factors or establish relationship between crashes and explanatory variables, such as traffic flows, type of traffic control, and highway geometric variables [7, 8 and 9].
In Nigeria, about 85% of the accounted causes of road accidents are believed to have been constituted by human factors [10]. Many researches carried out in Nigeria revealed that most accidents caused by human factors are the result of driving while drunk, drugs, inexperience or poor driving skills, health problems, psychological problems and temperament are also not left out. These have been shown in different ways by drivers. It is also noted that these human factors are the greatest contribution to the increasing surge of traffic accidents in Nigeria [1]. The attitude towards road traffic accidents includes such behavioral elements of the drivers as: sleeping while driving and tiredness, inadequate preparation for a journey, not been familiar with the highway signs, cutting corners, driving after taking excess alcohol, driving with bad eye sight especially in the night, ignorance of the use of seat belts, the incapability of handling unforeseen circumstances, wrong use of road signs and vehicle signaling, overtaking and incompetent maneuvering [11].

2.0 Methodology

2.1 Description of study area

The study area is Benin City, specifically Benin-Lagos Road. Benin City serves as the principal administrative and socio-economic center for both Oredo Local Government Area and Edo State in Nigeria. Benin City is a humid tropical urban settlement which comprises three Local Government Areas namely Egor, Ikpoba Okha and Oredo. It is located within latitudes $6^0 20'$N and $6^0 58'$N and longitudes $5^0 35'$E and $5^0 41'$E. It broadly occupies an area of approximately 112.552 sq km. This extensive coverage suggests spatial variability of weather and climatic elements. Benin City lies visibly in the Southern most corner of a dissected margin: a prominent topographical unit which lies north of the Niger Delta, west of the lower Niger Valley, and south of the Western Plains and Ranges [12]. Rainfall, temperature, wind and relative humidity are the most significant climatic elements in Benin City. The rainfall element strongly determines the occurrence of the wet and dry seasons in the study area. As observed during the assessment of the urban troposphere using sensitive rain gauges of the American origin, the rainfall amount, its intensity, duration as well as its distribution throughout the city are determined by the prevailing maritime winds, changing clouds, temperatures and circulating pressures. Two principal air masses prevail in the city. These are the tropical maritime and tropical continental. The tropical maritime air mass which is essentially humid, warm, moisture-borne, and widely resident in Benin City for almost twelve months, originates from the South Atlantic Zone. It causes rainfall which begins from the late January till its gradual subsidence in mid-November. The arrival of rainfall in the study area brings welcome relief to the urban residents from the prevailing moderately dry and cold wind periods which normally occur between late December and the end of January [12]. Heavy rainfall and the associated floods occur frequently in Benin City and have caused huge economic losses as well as social problems. The base map of the study area is presented in Figure 1.

2.2 Data collection

Two different types of data were employed in this study. The secondary data which include; road accident data was collected from Federal Road Safety Office in Benin City while the primary data which include; traffic volume and the geometric data was collected from the field. For a robust field data, a reconnaissance surveys was carried out at selected points of interest along the study area. For each selected point of interest, detailed information regarding accidents, traffic flow, geometric characteristics, traffic characteristics, road way condition, approach speed, lighting, among others were sourced.

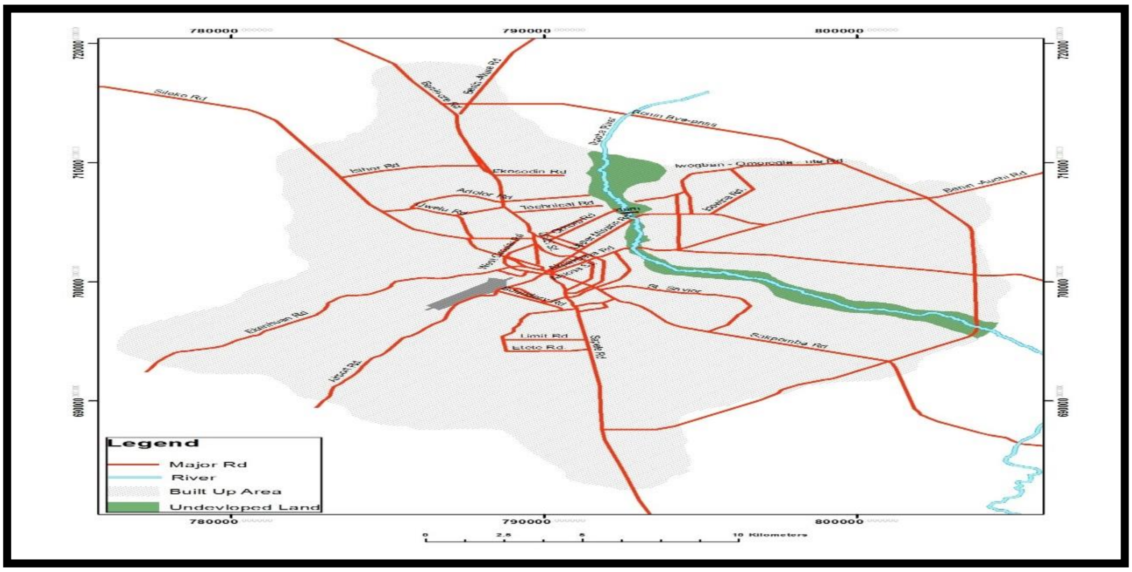

Figure 1: Base map of study area (Adopted from Google Earth)

2.3 Characterization of road geometry

To characterize the geometry of the road under study, the following information was collected; assessment of the presence of speed bumps, assessment of the presence of walkway, assessment of the presence of a shoulder, assessment of curves on the road and assessment of the presence of a median between the two carriage way. To assess the presence of the above information concerning the road, a reconnaissance survey was done and relevant road geometry were measured and recorded.

2.4 Statistical Analysis of Data

Some of the statistical analysis techniques employed to evaluate the quality of the data include; descriptive statistics, detection of outliers, test of reliability of data, test of normality and diagnostic statistics which include; heteroskedasticity test, autocorrelation test and variance inflation factor.

2.4.1 Descriptive Statistics

Descriptive statistics which include skewness, coefficient of variability and kurtosis measure was determine using Equation 1 and 2.

$$y = \frac{E(x-\mu)^3}{\sigma^3} \tag{1}$$

Where:

Y is the skewness, μ is the mean of x, σ is the standard deviation of x, E (t) is the expected value of the quantity x and SES is standard error of skewness. A value of kurtosis significantly greater than 0 indicates that the variable has longer tails than those for a normal distribution; less than 0 indicates that the distribution is flatter than a normal distribution. A kurtosis coefficient is considered significant if the absolute value of (KURTOSIS / SEK) is greater than 2. Mathematically, the kurtosis of a distribution is defined as [13].

$$k = \frac{E(x-\mu)^4}{\sigma^4} \tag{2}$$

Where:

K is the kurtosis, μ is the mean of x, σ is the standard deviation of x, E (t) is the expected value of the quantity x and SEK is the standard error of kurtosis. Coefficient of variation is the standard deviation divided by the sample mean. Mathematically, the coefficient of variation (CV) of a distribution is defined as:

$$CV = \left(\frac{\sigma}{\mu}\right) \tag{3}$$

Where: μ is the mean of x and σ is the standard deviation of x

2.4.2 Detection of outliers using the labeling rule

Although, the presence of outlier can be visualized using the histogram plot. In this study, the labeling rule method was employed to detect the presence of outliers. The labeling rule is the statistical method of detecting the presence of outliers in data sets using the 25th percentile (lower bound) and the 75th percentile (upper bound). The underlying mathematical equation based on the lower and the upper bound is presented as follows:

$$\textit{Lower Bound}\ \mathrm{Q}_1 - (2.2 \times (Q_3 - Q_1) \tag{4}$$

$$\textit{Upper}\ \mathrm{Bound}\ \mathrm{Q}_3 + (2.2 \times (Q_3 - Q_1) \tag{5}$$

Q is the lower bound, Q is the upper bound. At 0.05 degree of freedom, any data lower than Q_1 or greater than Q_3 will be considered an outlier and needed to be removed before further analysis [14].

2.4.3 Reliability Analysis of the Data

Reliability analysis was done to ascertain the fitness of the data for the selected analysis. The null and alternate hypothesis of reliability was formulated as follows;

H0: Data are reliable

H1: Data are not reliable

Using the Fisher's probability test (F-test), the analysis was conducted at p-value of 0.05. At p-value < 0.05, the null hypothesis was accepted and was concluded that the data are good and can be employed for further analysis.

2.4.4: Assessment of Normality

The Jarque-Bera test for normality is employed to ascertain whether the data follow a normal distribution. Mathematically, the Jarque-Bera test is define as follows"

$$JB = n[(\sqrt{b_1})^2 / 6 + (b_2 - 3)^2 / 24] \tag{7}$$

Where:

n is the sample size, √b is the sample skewness and b_2 is the kurtosis coefficient. The null hypothesis for the Jarque-Bera test is that the data are normally distributed while the alternate hypothesis is that the data does not come from a normal distribution. In which case;

H0 = Data follows a normally distributed

H1 = Data do not follow a normal distribution

In general, a large JB value indicates that the residuals are not normally distributed. A value of JB greater than 10 means that the null hypothesis has been rejected at the 5% significance level. In other words, the data do not come from a normal distribution. JB value of between (0-10) indicates that data is normally distributed.

2.5. Diagnostic Statistics

It is pertinent to note that standard error estimation and computation of t-statistics are appropriate in calculating the probability (p-value) by which you test the significance of the regression model. In the presence of heteroskedasticity, it is assumed that the overall standard error of regression and the t-statistics computed for each independent variable may not be completely adequate to estimate the resulting probability (p-value) of regression. In addition, the presence of serial correlation can lead to a number of issues, namely;

i. Make reported standard error and t-statistics to be invalid
ii. Coefficient may be biased, though not necessarily inconsistent

Based on this argument, selected diagnostic statistics were conducted to verify the statistical properties of the overall regression model. The selected diagnostic statistics include;

i. Heteroskedasticity test using Breusch-Pagan Godfrey
ii. Serial Correlation test using Breusch Godfrey
iii. Variance Inflation Factor (VIF)

2.5.1 Heteroskedasticity Test

Heteroskedasticity is a diagnostic test statistic use to diagnose the adequacy of the probability (p-value) calculated for each individual variable. Hence it is important to know whether there is or there isn't heteroskedasticity in our data. The null and alternate hypothesis of heteroskedasticity was formulated as follows;

H0 = Presence of homoscedasticity
H1 = Absence of heteroskedasticity
H0 = Absence of homoscedasticity
H1 = Presence of heteroskedasticity

For p-value < 0.05 you reject the null hypothesis of homoskedasticity and conclude that there is no heteroskedasticity. For p-value > 0.05 you accept the null hypothesis of homoskedasticity and conclude that there is the presence of heteroskedasticity.

2.5.2 Serial Correlation Test

Serial correlation is a common occurrence in time series data because the data is ordered (overtime). It is therefore not surprising that neighbouring error terms turn out to be correlated. Serial correlation violates the standard assumption of regression theory that error terms are uncorrelated. The null and alternate hypothesis of serial correlation was formulated as follows;

H0 = Absence of serial correlation
H1 = Presence of serial correlation

2.5.3 Variance inflation factor

Variance inflation factor (VIF) measures the correlation of the dependent variable with the independent variables. Ideal VIF is 1; VIF greater than 10 is cause for alarm showing the variables are uncorrelated due to multicollinearity.

3.0 Discussion of Results

Result of the geometry features of the road under study is presented in Table 1

Table 1: Geometric Characteristics of Ugbowo Benin-Ore Road

Chainage KM	Vertical curve %	Horizontal curve (M)	Super elevation %
11.5-13.0	12.4	2440.56	4.3
13.7-14.6	0	425.67	5.1
24.7-39.3	8.7	0	0
59.5-62.3	2.6	3642.47	1.5
72.8-73.8	0	2088.15	2.6
74.0-76.6	0	3290.26	1.2
78.0-81.0	1.3	5726.30	2.1
84.0-85.0	0	1022.96	0.4
86.0-87.0	0	5087.89	1.3
90.0-90.5	0	904.28	0.5

The geometry features of Ugbowo Benin-Ore road such as design speed, road width, and median Shoulders were 100km/hr, 10.5m, and 1.5m respectively. The length of the road is 94km and the AADT is 1850. In order to attain the primary goal of road transportation, road designers and their disciplines need to use different emerging technologies and techniques. Analysis of road geometric design consistency has been used widely to improve the safety of the roads. Geometric design consistency can be demarcated as how a driver expectation and the road performance match up (i.e. when the road with good constituency level matches a driver expectation, the road user is not amazed while driving along it). Design constituency corresponds to reliving the design speed with actual driving behaviour, which is expressed by the 85th percentile speed of passenger cars under free-flow conditions are prerequisites to accident free highways. Descriptive statistics of the data employed for the analysis is presented in Table 2.

Table 2: Descriptive statistics of accident data

Variable	Observations	Obs. with missing data	Obs. without missing data	Min.	Max.	Mean	Std. deviation
NAC	60	0	60	8.000	36.000	19.550	6.863
NPIV	60	0	60	55.000	365.000	175.917	74.651
NPIJ	60	0	60	23.000	180.000	67.150	31.338
NPK	60	0	60	2.000	33.000	13.083	7.158
NVI	60	0	60	11.000	76.000	37.050	14.611

NAC: Number of accident cases, NPIV: Number of persons involved, NPIJ: Number of persons injured, NPK: Number of persons killed and NVI: Number of vehicles involved.

Based on the results of Table 2, it was observed that the minimum case of accident from 2014 to 2018 was 8 while the maximum was 36 cases of accident. On the number of vehicles involved, the minimum number was 11 while the maximum number of vehicles was 76. On the number of persons involved, the minimum number was 55 while the maximum was 365. On the number of persons injured, the minimum was 23 while the maximum was 180. On the number of persons killed, the minimum was 2 while the maximum was 33.
The correlation matrix of regression which shows how the individual variables relates to the others is presented in Table 3.

Table 3: Correlation Matrix

Variables	NAC	NPIV	NPIJ	NPK	NVI
NAC	**1**	0.782	0.777	0.429	0.920
NPIV	0.782	**1**	0.741	0.376	0.862
NPIJ	0.777	0.741	**1**	0.394	0.773
NPK	0.429	0.376	0.394	**1**	0.367
NVI	0.920	0.862	0.773	0.367	**1**

Result of Table 3 revealed that the individual variables are strongly positively correlated with one another. For example, with a correlation coefficient of 0.777 it was concluded that the number of persons involved in an accident (NPIV) is strongly correlated with the number of persons injured (NPIJ). With a correlation coefficient of 0.429, the number of persons involved in an accident is poorly correlated with the number of persons killed (NPK). With a correlation coefficient of 0.920, it was concluded that the number of persons involved in an accident (NPIV) is most strongly correlated with the number of vehicles involved (NVI). For reliability analysis, it is important that analysis of variance is significant at the 5% confident limit. The computed analysis of variance is presented in Table 4.

Table 4: Analysis of variance

Source	DF	Sum of squares	Mean squares	F	Pr > F
Between subjects	59	176265.850	2987.557	3.081	**< 0.0001**
Within subjects	240	1298026.400	5408.443		
Between measures	4	1069161.733	267290.433	275.624	**< 0.0001**
Residual	236	228864.667	969.766		
Total	299	1474292.250	4930.743		

Computed against model Y=Mean(Y)

With probability p-value <0.0001 as observed in Table 4, it was concluded that the model is significant, hence the Cronbach alpha value for assessing reliability was calculated and presented in Table 5.

Table 5: Cronbach's alpha statistics

Cronbach's alpha	Standardized Cronbach's Alpha
0.675	0.900

For reliability, the Cronbach alpha value must be greater than 0.65. For standardized Cronbach alpha values of 0.900 as observed in Table 5, it was concluded that the accident data are reliable. Finally, the goodness of fit statistic of reliability were calculated and presented in Table 6

Table 6: Goodness of fit statistics of reliability

Variable	<Scale/deleted item> Mean	<Scale/deleted item> Variance	<Scale/deleted item> Correlation	<Scale/deleted item> R²	<Scale/deleted item> Cronbach's alpha	<Scale/deleted item> Guttman L6
NAC	293.200	13528.468	0.853	0.866	0.661	0.873
NPIV	136.833	2816.311	0.827	0.764	0.721	0.855
NPIJ	245.600	9271.702	0.776	0.753	0.787	0.851
NPK	299.667	14189.277	0.609	0.814	0.693	0.881
NVI	275.700	11874.214	0.895	0.901	0.786	0.857

Results of Table 6 further confirmed that the accident data are reliable with Guttman L6 coefficient of 0.851 to 0.881, coefficient of determination of 0.753 to 0.901, Cronbach alpha value of 0.661 to 0.787 and correlation coefficient of 0.609 to 0.895. Result of normality test based on Jarque-Bera approach is presented in Figures 2, 3, 4, 5 and 6 respectively.

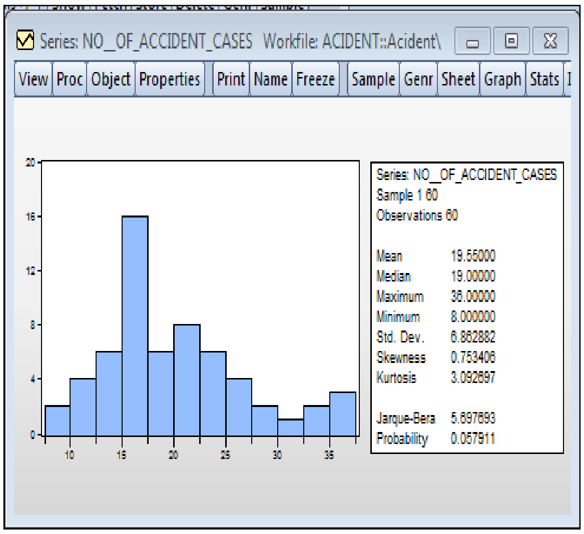

Figure 2: J-B normality test for NAC

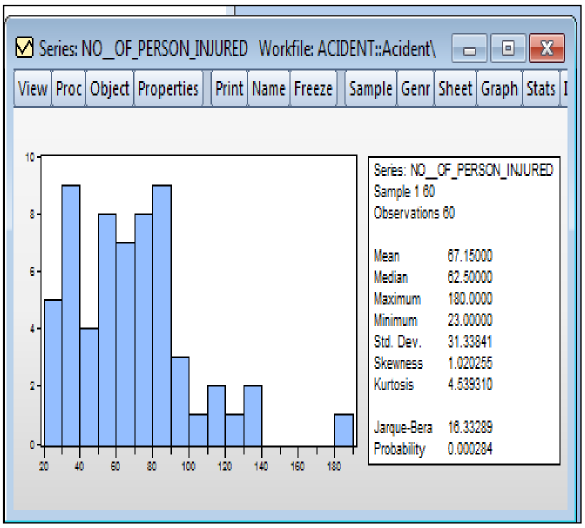

Figure 3: J-B normality test for NPIJ

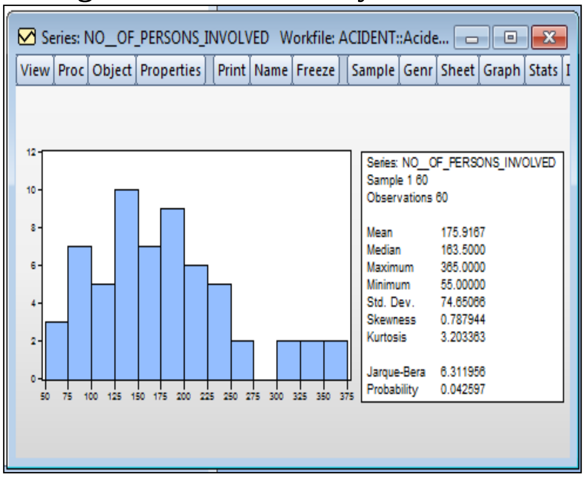

Figure 4: J-B normality test for NPIV

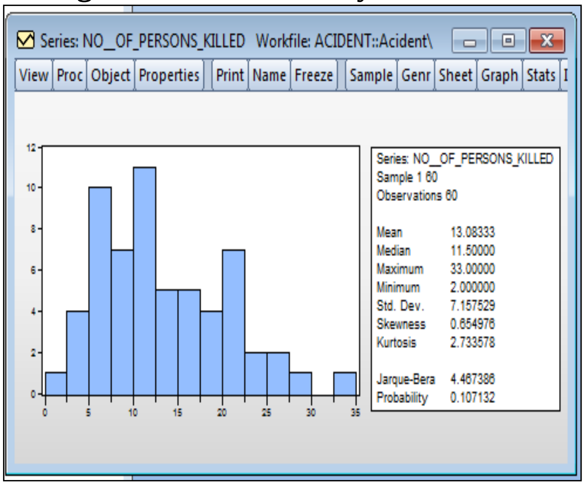

Figure 5: J-B normality test for NPK

In Figure 2, the computed Jarque-Bera test statistics was 5.697693 with a probability p-value of 0.057911. Although, the Jarque-Bera tests value is less than 10, the computed p-value was greater than 0.05. Hence, it was concluded that the data on number of accident cases is not normally distributed. For number of persons injured as observed in Figure 3, the calculated Jarque-Bera value was observed to be 16.33299 with a probability p-value of 0.002284. Although, the Jarque-Bera value is greater than 10, the calculated p-value is less than 0.05. Hence, it was concluded that the data on number of persons injured follows a normal distribution. For number of persons involved as observed in Figure 4, the calculated Jarque-Bera value was observed to be 6.311956 with a probability p-value of 0.042597. Since the Jarque-Bera value is less than 10, and the calculated p-value is less than 0.05, it was concluded that the data on number of persons involved follows a normal distribution. For number of persons killed as observed in Figure 5, the calculated Jarque-Bera value was observed to be 4.467386 with a probability p-value of 0.107132. Although, the Jarque-Bera value is less than 10, the calculated p-value is greater than 0.05. Hence, it was concluded that the data on number of persons killed did not obey normality.

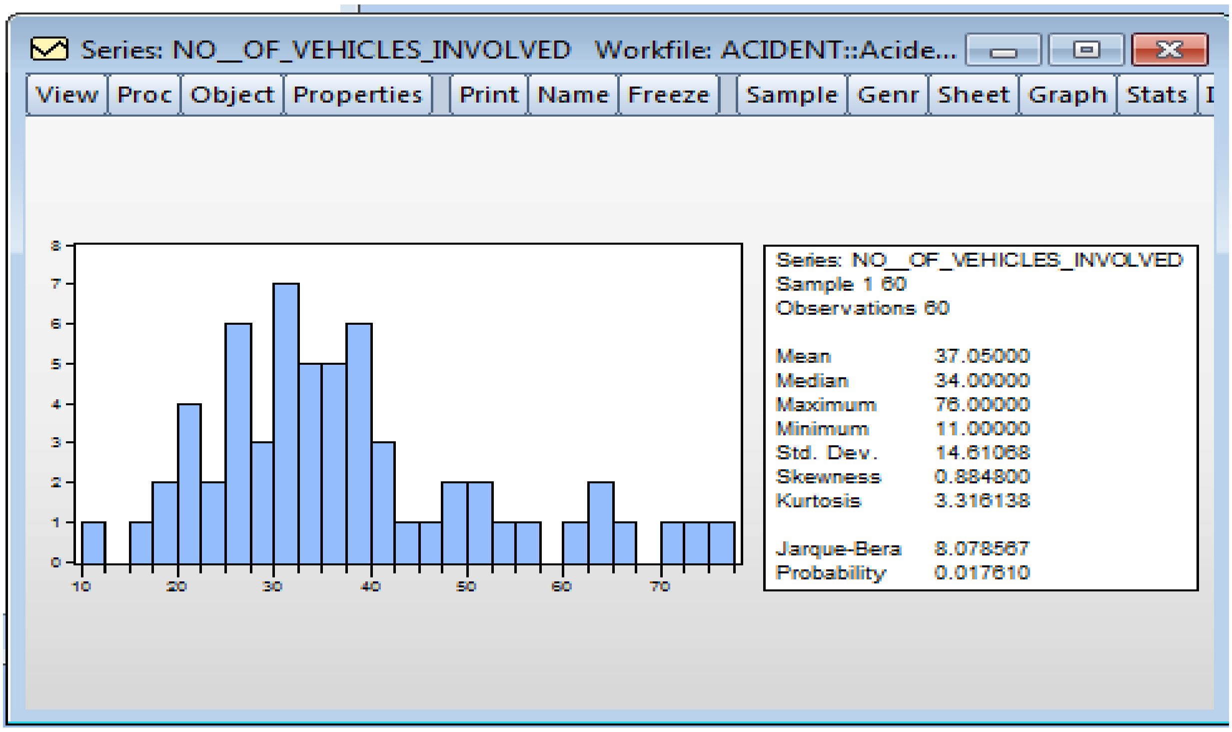

Figure 5: J-B normality test for NVIV

For number of vehicles involved as observed in Figure 6, the calculated Jarque-Bera value was observed to be 8.078567 with a probability p-value of 0.017610. Since the Jarque-Bera value is less than 10, and the calculated p-value is less than 0.05, it was concluded that the data on number of vehicles involved follows a normal distribution.
Result of the outlier detection test employed to evaluate the quality of the data based on the Dixon approach is presented in Table 7.

Table 7: Result of Dixon test for outliers

Parameters	R10 (Observed value)	R10 (Critical value)	p-value (Two-tailed	Alpha
NAC	0.036	0.244	0.628	0.05
NPIV	0.039	0.244	0.673	0.05
NPIJ	0.274	0.244	0.725	0.05
NPK	0.129	0.244	0.609	0.05
NVI	0.077	0.244	0.84	0.05

As the computed p-value is greater than the significance level of 0.05, one cannot reject the null hypothesis H0. Hence, we accept the null hypothesis and concluded that there are no outliers in the accident data. Results of the diagnostic statistics which include heteroskedasticity test, autocorrelation test and variance inflation factor is presented in Tables 8, 9 and 10 respectively

Table 8: Result of heteroskedasticity test

Equation: EQ1 Workfile: ACCIDENT DATA SPSS::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	1.215502	Prob. F(4,55)	0.3148
Obs*R-squared	4.873214	Prob. Chi-Square(4)	0.3006
Scaled explained SS	4.568656	Prob. Chi-Square(4)	0.3345

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 07/06/21 Time: 10:59
Sample: 1 60
Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.723035	3.488870	0.207241	0.8366
NO__OF_VEHICLES_INVOLVED	0.274056	0.177194	1.546645	0.1277
NO__OF_PERSON_INJURED	0.018019	0.063115	0.285495	0.7763
NO__OF_PERSONS_INVOLVED	-0.030136	0.032840	-0.917647	0.3628
NO__OF_PERSONS_KILLED	-0.045960	0.184771	-0.248739	0.8045

R-squared	0.081220	Mean dependent var	6.184110
Adjusted R-squared	0.014400	S.D. dependent var	9.315730
S.E. of regression	9.248414	Akaike info criterion	7.366436
Sum squared resid	4704.324	Schwarz criterion	7.540965
Log likelihood	-215.9931	Hannan-Quinn criter.	7.434704
F-statistic	1.215502	Durbin-Watson stat	1.614077
Prob(F-statistic)	0.314771		

From the result of Table 8, it was observed that;

i. The calculated (p-value) based on the F-statistics is 0.3148
ii. The calculated (p-value) based on langrange multiplier (LM) is 0.3006

Since the computed (p-value) based on F-statistics and langrange multiplier is greater than 0.05 (P > 0.05), we accept the null hypothesis of homoskedasticity and conclude that there is heteroskedasticity in the data. The implication is that linear regression may not be the best model to assess the relationship between the accident data.

Table 9: Result of Serial Correlation Test

Equation: EQ1 Workfile: ACCIDENT DATA SPSS::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.020058	Prob. F(2,53)	0.9801
Obs*R-squared	0.045379	Prob. Chi-Square(2)	0.9776

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 07/06/21 Time: 11:09
Sample: 1 60
Included observations: 60
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.020344	1.015356	-0.020037	0.9841
NO__OF_VEHICLES_INVOLVED	0.003054	0.052948	0.057671	0.9542
NO__OF_PERSON_INJURED	0.000188	0.018904	0.009944	0.9921
NO__OF_PERSONS_INVOLVED	-0.000542	0.009839	-0.055040	0.9563
NO__OF_PERSONS_KILLED	-0.000819	0.053348	-0.015344	0.9878
RESID(-1)	-0.024849	0.141825	-0.175212	0.8616
RESID(-2)	0.014926	0.151505	0.098517	0.9219

R-squared	0.000756	Mean dependent var	2.04E-16
Adjusted R-squared	-0.112366	S.D. dependent var	2.507773
S.E. of regression	2.644917	Akaike info criterion	4.892437
Sum squared resid	370.7660	Schwarz criterion	5.136777
Log likelihood	-139.7731	Hannan-Quinn criter.	4.988012
F-statistic	0.006686	Durbin-Watson stat	1.988191
Prob(F-statistic)	0.999999		

From the result of Table 9, it was observed that;

i. The calculated (p-value) based on the F-statistics is 0.9801
ii. The calculated (p-value) based on langrange multiplier (LM) is 0.9776

Since the computed (p-value) based on F-statistics and langrange multiplier is greater than 0.05 (P > 0.05), we accept the null hypothesis of serial correlation and concluded that there is no serial correlation in the data

Table 10: Calculated variance inflation factors

Equation: EQ1 Workfile: ACCIDENT DATA SPSS::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Variance Inflation Factors
Date: 07/06/21 Time: 11:14
Sample: 1 60
Included observations: 60

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	0.960066	8.538595	NA
NO__OF_VEHICLES...	0.002476	34.85694	4.623334
NO__OF_PERSON_I...	0.000314	15.29861	2.698573
NO__OF_PERSONS...	8.51E-05	27.55772	4.145661
NO__OF_PERSONS...	0.002693	5.305859	1.206455

Since the computed variance inflation factors (centered VIF) for the selected independent variables are less than 10, it was concluded that the variables are well correlated with the dependent variable, hence absence of multicollinearity.

4.0 Conclusion

In this study, selected statistical approaches, namely; descriptive analysis, reliability analysis, test of normality, outlier detection including diagnostic statistics such as heteroskedasticity test, autocorrelation test and variance inflation factor have been employed to assess the quality of five (5) years monthly accident data for Benin-Lagos road, Benin City, Edo State Nigeria. Results of the analysis have shown that; the accident data collected are reliable with Cronbach Alpha Coefficient value of > 0.80. In addition, it was observed that; there is no multicollinearity in the data since the centered variance inflation factor (CVIF) values calculated are less than 10

References

[1] Odumosu, A. O. (2005); Demeanour Transposition as Strategy for Traffic Accident Reduction in Nigeria: Case Study of Niger State, Nigeria, Proc. 13th International Conference on Road Safety on Four Continents, Warsaw, Poland, p. 18
[2] World Health Organization (WHO), (2018): Global status report on road accident and safety, pp: 23-56
[3] Ozgan, E. and Demirci, R. (2008); Neural Networks-based Modeling of Traffic Accidents in Interurban Rural Highways, Duzce Sampling; Journal of Applied Science, vol. 8, pp. 146-151
[4] Ohakwe, J., Iwueze, I. S., Chikezie, D. C. (2011); Analysis of Road Traffic Accidents in Nigeria: A Case study of Obinze/Nekede/Iheagwa Road in Imo State, southeastern, Nigeria. Asian Journal of applied Sciences, 4(2): pp. 166-175
[5] Oladehinde, M. K, Adeoye, A. O, Adegbehingbe, B. O, and Onakoya A. O. (2007) Visual Functions of Commercial Drivers in Relation to Road Accidents in Nigeria. Indian Journal of Occupation and Environmental Mediation, Vol. 11, pp. 71-75

[6] Helai, H. and Mohamed, A. A. (2010); Multilevel data and Bayesian analysis in traffic safety. Accident Analysis and Prevention, Retrieved on October 18, 2010, from www.elsevier.com/locate/aap, vol. 42, pp: 1556–1565

[7] Gwynn, D.W. (1967); Relationship between Road Accident and Hourly Volumes, Traffic Quarterly, Vol. 14(1), pp: 407-418

[8] Ghani, A.R.A., Daniel, B.D., Yusof, I., Zamhari, K.A., Abdullah, M.E., Omar, N. and Termida, N.A. (2008); Traffic Engineering and Safety Module. 2nd Ed. Johor: Penerbit UTHM, pp" 67-78

[9] Fajaruddin, M., Ismail, Y., Ismail, R., Aziz, A. S., and Esnizah, M. S. (2008); Black spot Study and Accident Prediction Model Using Multiple Liner Regression. Proceedings of the First International Conference on Construction in Developing Countries; Karachi, Pakistan: Universiti Tun Hussein Onn Malaysia, vol. 1, pp: 345-366

[10] Osime, O. C., Ehikhamenor, E., Oludiran, O. O., Iribhogbe, P. E., Ighedosa, S. E. and Elusoji, S. O. (2006) Road Traffic Accident Trends in Nigeria. International Journal of Injury Control and Safety Promotion, 13(4), pp. 251-253

[11] Okhakhu, P.A. (2010). The Significance of Climatic Elements in Planning the Urban Environment of Benin City, Nigeria, Unpublished PhD Thesis, Ekpoma-Nigeria: Department of Geography and Regional Planning, Ambrose Alli University.

[12] Tatiana, T., Dora, S., Miltiadis, C and Andreas, G. (2014), Maximizing Accuracy and Efficiency of Traffic Accident Prediction Combining Information Mining with Computational Intelligence Approaches and Decision Trees, Jaiscr, Vol.4 (1), Pp.31-42.

[13] Levi, D. B.; Julie, E. K.; Olsen, J. R.; Pulwarty, R. S.; Raff, D. A.; Turnipseed, D. P.; Webb, R. S and Kathleen D. W (2009); Climate Change and Water Resources Management: A Federal Perspective, circular 1331, pp: 1 – 72