Journal of NAMP 69, 1 (2025) 1-18

The Nigerian Association of Mathematical Physics

Journal homepage: https://nampjournals.org.ng

ANALYSIS OF THE CAUSES OF ROAD TRAFFIC ACCIDENT IN NIGERIA: A MULTIVARIATE ANALTYTIC APPROACH

***Adetona, B.O.[1], Mojeed, S.A.[2], Adetona, M.O.[3], Agwuegbo, S.O.N.[4], Sowola, K.T.[5]**

[1,2,3,4] *Department of Statistics, Federal University of Agriculture, Abeokuta, Nigeria*

[5] *Department of Statistics, University of Ilorin, Nigeria*

ARTICLE INFO

Article history:
Received xxxxx
Revised xxxxx
Accepted xxxxx
Available online xxxxx

Keywords:
RTA,
Dendrogram,
Eigen Value,
Principal Component,
Factors Loadings,
Uniqueness,
Communality,

ABSTRACT

This study investigates the root causes of road traffic crash on the Nigerian highway using multivariate analytic tools. The dataset for the study was obtained from NBS website, it consists of the number of road traffic crashes recorded as a result of eighteen conventional events across the 36 states and the Federal Capital Territory (FCT) of the federation. R-software was used for the analysis. Correlation matrix of the eighteen variables was obtain and used to conduct a principal component analysis. Factor analysis was further carried out on the data, the result showed that not less than nine factors is required to explain the variation in the data. Factors identified in this study describe the root cause of RTA on the Nigerian highway. They include: motorist' exhaustion level, traffic rules violation, greed of commercial motorist to maximize profit, structural integrity of tyres, bad roads, recklessness of motorists on the highway and others

1. Introduction

Road traffic Accidents (RTA) is regarded as one of the major causes of mortality and morbidity around the world and it has been discovered that low and middle-income countries are the most affected [9]. According to [21], millions of lives are lost annually due to injuries from road traffic crashes. Africa as a continent, accounts for a significant percentage of global road fatalities despite having few fractions of the world's vehicle. Hence Africa cannot be exempted from the global road safety crisis.

In Nigeria, road crashes are identified as one of the leading causes of death especially among the younger age-group of the population. Between 1960 and 1988, the Nigerian police force was

*Corresponding author: Adetona, B.O.

E-mail address: sowolabo@funaab.edu.ng

https://doi.org/10.60787/jnamp.vol69no1.462

saddled with the responsibility to collect and collate road accident data. However, since the formation of Federal Road Safety Commission (now Federal Road Safety Corps (FRSC)) in 1988, it became a statutory responsibility of the corps to collect and collate road traffic accident data nationwide. According to [10], road traffic crashes occurrences in Nigeria always result to either loss of lives or injuries that poise temporal or permanent disability.

Vehicular traffic can be thought of, when there is at least a vehicle (mechanical) to be driven, a driver (human) to drive and a road (environment) to drive on, hence a crash can only occur when there is a deficient in the inter relationship in any of these three factors [11]. Human error contributes larger percentage to the cause of traffic crashes in developing countries ([8]; [18]; [3]). The most important human-related factor contributing to road traffic accidents in Nigeria is the attitude of the driver to driving code and etiquettes [4]. Human-related issues include speed violation, dangerous driving, sleeping on steering, driving under alcohol/drug influence, use of phone while driving and traffic light violation. Vehicular factors that could result to road traffic accident include vehicle design, brake system, vehicle tyres, vehicle light and engine. Every vehicle is designed for specific maximum load, so it is no surprise that when a vehicle is subject to stress above the design specification, accelerated wear and tear are banned to be experienced by such vehicle [7]. The brake subsystem works jointly with accelerator and synchronizes the speed of vehicles, any malfunction of the brake system is to be taken as a potential source of accident. When tyres are not in good form and vehicle light are malfunctioning, accident can occur. Apart from the two causes road traffic accident mentioned above, environmental factors can also contribute greatly to the rate of road accidents in Nigeria. Some of these factors are unfavorable weather, bad road, and flood. Deficiencies of Nigerian roads are due to inadequate road design specification and poor maintenance culture [1]. Other significant factors include the frequency of potholes on the roads, the indiscriminate location of police check point and the reluctance of the appropriate authorities to continually improve on the condition of the roads.

[17] hinted that road traffic crash is presently the eleventh leading cause of death in the globe and projected that it may rise to seventh position by the year 2030. By logical reasoning one can attribute the increased road traffic accident globally to population explosion which subsequently could warrant an increased in level of motorization to meet up to the transport demand of people. Although causative factor of road traffic accident differs from country to country. [19], attributed the road traffic accident incidences in the US to driving under influence of alcohol. [15] identified that most road traffic accidents in in India is majorly caused by ignorance of motorist about preventive measures put in place to avert road traffic accident on Indian Highway. Nigeria recorded her first road traffic accident in in the year 1906, ever since then, the report of road traffic accidents on the Nigeria highways has been consistent and always attributed to different causative factors. This study intends to reveal those factors that majorly contribute to the crashes experienced on the Nigerian highways through the use of multivariate analysis techniques. Knowledge from this study will serve as guide to policy makers in designing policies that will tackle the menace of road traffic crashes in Nigeria. It will also serve as guide in earmarking budgets for the road safety commission for effective enforcement of road traffic regulations on the highway.

2. RESEARCH METHODOLOGY

This study aimed to identify those factors out of numerous, that are peculiar in causing road accident on the Nigerian highway. The study adopted the multivariate analytic tools which include Principal Component Analysis (PCA) targeted for data reduction and Factor Analysis (FA) for substantive interpretation.

The dataset for the study was collated by the Federal Road Safety Corp (FRSC) and made available for download on NBS website. It consists of the number of road traffic crashes recorded as a result of eighteen conventional causes across the 36 states and the Federal Capital Territory (FCT) of the Federation. The data is a quarterly data spanning through the year 2021 to 2023.

2.1 Multivariate Analysis

In multivariate analysis, the goal is to identify patterns, correlations and relationships between multiple variables. The information obtained thus can be used to identify underlying structures and dimensions. Multivariate data arises whenever a researcher seeks to understand a social or physical phenomenon by collating data about the various contributing factors (variables) likely to influence or affect the phenomenon of interest. The number of variables " p " to be considered in regarded must be greater than 2, i.e. $p > 2$. Observation (values) are for each of the variable are the recorded for " n " distinct object (an object in this case is the experimental unit).

Suppose $X_1, X_2, ...X_p$ represents $p-\text{variables}$ involved in an experiment and $x_1, x_2, ...x_p$ denote the observation (input) for each variable, then for n-objects the dataset is shown below:

	X_1	X_2	.	.	.	X_k	.	.	.	X_p
Object 1:	x_{11}	x_{12}				x_{1k}				x_{1p}
Object 2:	x_{21}	x_{22}				x_{2k}				x_{2p}
.	.	.				.				.
.	.	.				.				.
.	.	.				.				.
Object j:	x_{j1}	x_{j2}	.	.	.	x_{jk}	.	.	.	x_{jp}
.	.	.				.	.			.
.	.	.				.		.		.
.	.	.				.			.	.
Object n:	x_{n1}	x_{n2}	.	.	.	x_{nk}	.	.	.	x_{np}

Where $x_{jk} = j^{th}$ observation for the k^{th} variable. The dataset displayed above can be written in matrix form. Suppose X denotes the data in matrix form

$$X = \begin{pmatrix} x_{11} & x_{12} & . & . & . & x_{1k} & . & . & . & x_{1p} \\ x_{21} & x_{22} & . & . & . & x_{2k} & . & . & . & x_{2p} \\ . & . & & & & . & & & & . \\ . & . & & & & . & & & & . \\ . & . & & & & . & & & & . \\ x_{j1} & x_{j2} & . & . & . & x_{jk} & . & . & . & x_{jp} \\ . & . & & & & . & . & & & . \\ . & . & & & & . & & . & & . \\ . & . & & & & . & & & . & . \\ x_{n1} & x_{n2} & . & . & . & x_{nk} & . & . & . & x_{np} \end{pmatrix} \quad (1)$$

From equation, X is a $p \times n$ matrix in which the elements of X consists of all the observation recorded for the variables $X_1, X_2, ..., X_p$.

In vector form X can be written as

$$X = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_p \end{bmatrix} \quad (2)$$

From equation (2) above, X in this case is called a p-dimension vector because it consists of p-random variables. Hence X is referred to as random vector.

2.1 Random Vector

A random vector X is a vector whose elements are random variables defined on a probability space (Ω, F, P) such that

$$X : \Omega \rightarrow \square^{n} \qquad \forall \quad n \in \square \tag{3}$$

Use of random vectors in multivariate analysis makes it easy to incorporate some basic facts from linear algebraic thereby making computations more compact [12].

2.2.1 Orthogonal and Orthonormal vectors

A random vector X is said to be orthogonal if its minor product $(X^{'}X)$ yields a diagonal matrix, i.e.

$$X^{'}X = D \tag{4}$$

Where:

$$D = \begin{pmatrix} d_1 & 0 & . & . & . & 0 & . & . & . & 0 \\ 0 & d_2 & . & . & . & 0 & . & . & . & 0 \\ . & . & & & & . & & & & . \\ . & . & & & & . & & & & . \\ . & . & & & & . & & & & . \\ 0 & 0 & . & . & . & d_k & . & . & . & 0 \\ . & . & & & & . & & & & \\ . & . & & & & . & & & & \\ . & . & & & & . & & & & \\ 0 & 0 & . & . & . & 0 & . & . & . & d_n \end{pmatrix} \tag{5}$$

And $$\sum_{i=1}^{n} d_i = trace(D) \tag{6}$$

A vector X is said to be orthonormal if its minor product yields an identity matrix. i.e

$$\boldsymbol{X}^{'}\boldsymbol{X} = \mathbf{I}_{n} \tag{7}$$

Where:

$$\mathrm{I}_n = \begin{pmatrix} 1 & 0 & . & . & . & 0 \\ 0 & 1 & . & . & . & 0 \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ 0 & 0 & . & . & . & 1 \end{pmatrix} \tag{8}$$

An Orthonormal vector is an orthogonal vector with a unit length. It is a normalized version of an orthogonal vector [2].

The concepts of orthogonality and orthonormality find relevance in some statistical procedures which include Principle Component Analysis (PCA).

2.2.2 Norm of random vector

For a random vector $X = (X_1, X_2, ..., X_p)^T$, the norm is denoted by $\|X\|$, then the k-norm of X is defined as :

$$\|X\|_k = \left(\sum_{i=1}^{p} X_i^k\right)^{1/k}$$
$$= \sqrt[k]{X_1^k + X_2^k + ... + X_p^k} \qquad (9)$$

Hence 2 -norm of the random vector X can be written as

$$\|X\|_2 = \sqrt{X_1^2 + X_2^2 + ... + X_p^2}$$
$$= \sqrt{\sum_{i=1}^{p} X_i^2} \qquad (10)$$

In terms of matrix X

$$\|X\|_2 = \sqrt{(X^{'}X)}$$
$$= (X^{'}X)^{1/2} \qquad (11)$$

From equation (10), it can be seen that 2 -norm is equivalent to the "square root of sum of squares. The 2 -norm vector is therefore referred to as the Euclidean norm which is equivalent to the length of vector X .

2.2.3 Normalized Vector

Given a random vector X , the normalized vector of X is a vector in the same direction as X but with a unit length [20]. Suppose $\tilde{X}$ is the normalized vector of X . Then $\tilde{X}$ is defined as

$$\tilde{X} = \frac{X}{\|X\|_2} \qquad (12)$$

Where $\|X\|_2$ is the Euclidean norm of the random vector X .

$$\tilde{X} = \frac{X}{(X^{'}X)^{1/2}} \qquad (13)$$

2.3 Covariance and Correlation Matrix

For a random vector X , which consists of p -random variables $X_1, X_2, ..., X_p$, the mean vector is defined as:

$$E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ . \\ . \\ . \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ . \\ \mu_p \end{bmatrix} = \mu \qquad (14)$$

The covariance of random variable X_i and X_j is:

$$\text{cov}(X_i X_j) = E(X_i X_j) - E(X_i)E(X_j) \qquad (15)$$

Where:

$$E(X_i) = \mu_i \quad \text{and} \quad E(X_j) = \mu_j$$

$$\therefore \quad COV(X_i X_j) = E(X_i X_j) - \mu_i \mu_j$$
$$= \sigma_{ij} \qquad (16)$$

Hence the covariance matrix will be a square matrix containing the covariance between each pair of random X_i and X_j element of random vector X . A covariance matrix is be denoted by Σ , such that

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & . & . & . & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & . & . & . & \sigma_{2p} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ \sigma_{p1} & \sigma_{p2} & . & . & . & \sigma_{pp} \end{pmatrix} \quad (17)$$

Where Σ is a $p \times p$ matrix whose diagonal elements are variances. The covariance matrix of a random vector X can be written as:

$$Cov(X) = E[(X - E(X))(X - E(X))^{\mathrm{T}}]$$
$$= E\left[(X - \mu)(X - \mu)^T\right] \quad (18)$$

Where μ is the mean vector.

$$\Rightarrow Cov(X) = E(XX^T) - \mu_x \mu_x^{\mathrm{T}} \quad (19)$$

The correlation matrix of a random vector X is defined as the normalized form of a covariance matrix. Correlation matrix of random vector X is denoted by ρ_X, mathematically, the correlation of a random vector X with covariance matrix $COV(X)$ is:

$$\rho_X = \frac{COV(X)}{\|X\|} \quad (20)$$

Where $\|X\|$ is the norm of random vector X

Recall that $\|X\| = (X^{'}X)^{1/2}$

$$\therefore \quad \rho_x = \frac{COV(X)}{(X^{'}X)^{1/2}} \quad (21)$$

Where $0 \le \rho_x \le 1$

Both the covariance matrix and correlation serve as basis for some multivariate procedures such as the principal component analysis, PCA).

2.4 Eigen values and Eigen vector

Let A be a $k \times k$ square matrix and I be the $k \times k$ identity matrix, then the scalars $\lambda_1, \lambda_2, \ldots, \lambda_k$ satisfying the polynomial equation:

$$|A - \lambda \mathrm{I}| = 0 \quad (22)$$

are called the characteristic roots (eigenvalues or latent roots) of a matrix A. While equation (22) $|A - \lambda \mathrm{I}| = 0$ is called the characteristic equation.

Also, if there exists a non-zero vector X such that

$$AX = \lambda X \quad (23)$$

Then X is said to be an eigen vector (characteristic vector) of the matrix A associated with the eigenvalue λ.

2.5 Principal Component Analysis

According to [14], Principal component analysis is concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables. The purpose of a principal component analysis is to transform the matrix X of p -variates, which may be correlated, into another matrix Y of p -uncorrelated hypothetical variates which decrease in variance from first to last. Although p components are required to reproduce the total system variability, but most times, this variability can be accounted for by a small number k of new set of variables called the Principal Components. Principal Components are linear combination of the original variables that maximally explain the variance in the dataset [6]. This implies that there is as much information in the k components as there is in the original p variables. The k principal components can then replace the initial p variables and the original data set, consisting of n measurements.

Algebraically, principal components are particular linear combinations of the p random variables $X_1, X_2, ..., X_p$. These linear combinations represent a new coordinate system obtained by rotating the original system with $X_1, X_2, ..., X_p$ as the coordinate axes. The new axes represent the directions with maximum variability and provide a simpler and more parsimonious description of the covariance structure. Principal component depends solely on the covariance matrix $\sum$ (or the correlation matrix ρ) of $X^{T} = \left[X_1, X_2, ..., X_p\right]$. The formation of principal components does not necessary require a multivariate normal assumption [14].

Suppose the random vector $X^{'} = \left[X_1, X_2, ..., X_p\right]$has covariance matrix Σ , with eigenvalue–eigenvector pairs $(\lambda_1, e_1), (\lambda_2, e_2), ..., (\lambda_p, e_p)$, $\lambda_1 > \lambda_2 > . \quad . \quad . > \lambda_p > 0$, then the ith principal component is given by:

$$Y_i = e_i^{/} X_1 = e_{i1}X_1 + e_{i2}X_2 + ... + e_{ip}X_p, \qquad i = 1, 2, ...p \tag{24}$$

Where the collection $\{Y_i,\ i = 1, 2, ..., p\}$ are uncorrelated linear combinations with maximum variability measure. The principal components $\{Y_i\}$ are defined such that

$$\operatorname{var}(Y_i) = e_i^{'} \sum e_i = \lambda_i, \qquad i = 1, 2, ..., p \tag{25}$$

$$Cov(Y_i, Y_k) = e_i^{'} \sum e_k = 0, \qquad \forall\ i \neq k \tag{26}$$

2.5.1 Total variance of principal components

Suppose $\Sigma = P\Lambda P^{/}$ (defined in terms of the eigenvalues and eigenvectors)
Where:
Λ is the diagonal matrix of the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_p$ and $P = \left(e_1, e_2, ..., e_p\right)$, so that P is orthonormal (i.e. $PP^{/} = P^{/}P = I$). Then the trace of Σ is

$$\begin{aligned} tr(\Sigma) &= tr(P\Lambda P^{/}) \\ &= tr(PP^{/}\Lambda) \\ &= tr(\Lambda) \\ &= \lambda_1 + \lambda_2 + ... + \lambda_p \end{aligned} \tag{27}$$

Recall that the elements in the leading diagonal of Σ are $\sigma_1^{2}, \sigma_2^{2}, ..., \sigma_p^{2}$, then

$$tr\left(\Sigma\right) = \sigma_1^{2}, \sigma_2^{2}, ..., \sigma_p^{2}$$

$$tr\left(\Sigma\right) = Var(X_1) + Var(X_2) + ... + Var(X_p) = \sum_{i=1}^{p} Var(X_i) \tag{28}$$

Equation (27) becomes

$$\sum_{i=1}^{p} Var(X_i) = \lambda_1 + \lambda_2 + ... + \lambda_p$$

$$\Rightarrow \sum_{i=1}^{p} Var(X_i) = \sum_{i=1}^{p} \lambda_i$$

But from equation (25) $Var(Y_i) = \lambda_i$

$$\therefore \sum_{i=1}^{p} Var(X_i) = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} Var(Y_i)$$

$$\Rightarrow \sum_{i=1}^{p} Var(Y_i) = \sum_{i=1}^{p} \lambda_i \quad (29)$$

Thus, the proportion of total variance explained by the k^{th} principal component is:

$$\frac{\lambda_k}{\sum_{i=1}^{p} \lambda_i} \qquad \forall \; k = 1, 2, ..., p \quad (30)$$

If the first few principal components can explain up to 80% variation in the dataset, then these component can replace the original p-variables without much of loss information.

The components of the eigenvector $e_i^{/} = \left(e_{i1}, ..., e_{ik}, ..., e_{ip}\right)$ are very vital in principal component analysis as they are important in describing the loadings of variables in a principal component (PC). Loadings are correlation coefficients between X_k and Y_i. They are purposefully used in describing the relative importance of the various variables is in a given PC.

2.6 Factor Analysis

Factor analysis can be considered an extension of principal component analysis, both can be viewed as attempts to approximate the covariance matrix Σ. However, the approximation based on the factor analysis model is more elaborate. According to [14], Factor analysis can be utilized to examine the underlying patterns or relationships for a large number of variables and to determine whether the information can be condensed or summarized in a smaller set of latent variables which are referred to as factors.

For the random vector $X^{'} = \left(X_1, X_2, ..., X_p\right)$ with a mean vector μ as defined in equation (14), with variance-covariance matrix Σ, the factor model is defined as

$$X - \mu = LF + \varepsilon \quad (31)$$

Where $L = (p \times m)$ matrix of factor loadings

$F = m-$dimesional vector of the latent variable

$\varepsilon = p-$dimesional vector of latent error terms

The factor model described in (31) is said to be orthogonal if:

(i) ε_i and F_j are independent $\forall$ pairs (i, j)

(ii) $E\left(F\right) = 0$ and $Cov\left(F\right) = E(FF') = I_m$

(iii) $E\left(\varepsilon\right) = 0$ and $Cov\left(\varepsilon\right) = \psi$

Where ψ is a diagonal matrix of order p.

In matrix form the covariance structure of an orthogonal factor model can be define as:

$$Cov\left(X\right) = \Sigma = E(X-\mu)(X-\mu)^T$$

$$= E\left[(LF+\varepsilon)(LF+\varepsilon)^T\right]$$

$$= E\left[L(FF^{'})L^{'} + L(\varepsilon^{T}F) + L^{T}(F^{T}\varepsilon) + (\varepsilon\varepsilon^{T})\right]$$

$$= LE[FF^{'}]L^{'} + E[\varepsilon F^{'}]L^{'} + LE[F\varepsilon^{'}] + E[\varepsilon^{'}\varepsilon]$$

Recall that $E(FF') = I_m$; ε_i and F_j are independent, this implies that $E[\varepsilon F^{'}] =$ zero matrix and $E[F\varepsilon^{'}]$ = zero matrix

$$\therefore \Sigma = LL^{'} + \psi \tag{32}$$

Where ψ is a $p \times p$ diagonal matrix

2.6.1. Number of factors required

According to [5], the number of sufficient factors m required to explain the covariance structure can be determined using a likelihood-ratio test provided the variables $\{x_1, x_2, ..., x_p\}$ are Gaussian and the factor model has been fitted using maximum likelihood method. Other means of determining the number of factors is by applying some rule of thumbs (Kaiser's rule) which include: selecting the eigenvalues whose values are larger than 1.0 and a visual inspection from the scree plot. However, these rule of thumbs is applicable when the correlation matrix used in place of the covariance matrix when fitting the factor model [5]

2.6.2. Proportion of variance explained by the factors

From equation (32),

$$\Sigma = LL^{'} + \psi$$

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & . & . & . & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & . & . & . & \sigma_{2p} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ \sigma_{p1} & \sigma_{p2} & . & . & . & \sigma_{pp} \end{pmatrix} = \begin{pmatrix} \ell_{11} & \ell_{12} & . & . & . & \ell_{1m} \\ \ell_{21} & \sigma_{22} & . & . & . & \ell_{2m} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ \ell_{p1} & \ell_{p2} & . & . & . & \ell_{pm} \end{pmatrix} \begin{pmatrix} \ell_{11} & \ell_{21} & . & . & . & \ell_{p1} \\ \ell_{12} & \sigma_{22} & . & . & . & \ell_{p2} \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ \ell_{1m} & \ell_{2m} & . & . & . & \ell_{pm} \end{pmatrix} + \begin{bmatrix} \Psi_1 & 0 & . & . & . & 0 \\ 0 & \Psi_2 & . & . & . & 0 \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ 0 & 0 & . & . & . & \Psi_p \end{bmatrix}$$

$$\begin{aligned} \Rightarrow \sigma_{11} &= \ell_{11}^2 + \ell_{12}^2 + \ldots + \ell_{1m}^2 + \psi_1 \\ \sigma_{22} &= \ell_{21}^2 + \ell_{22}^2 + \ldots + \ell_{2m}^2 + \psi_2 \\ &\vdots \\ \sigma_{pp} &= \ell_{p1}^2 + \ell_{p2}^2 + \ldots + \ell_{pm}^2 + \psi_p \end{aligned} \tag{33}$$

In summary equation 33 can be written as:

$$\sigma_{ii} = \ell_{i1}^2 + \ell_{i2}^2 + \ldots + \ell_{im}^2 + \psi_i \qquad \forall\ i = 1, 2, ..., p \tag{34}$$

Where:

$\sigma_{ii} = Var(x_i)$

$\ell_{i1}^2 + \ell_{i2}^2 + \ldots + \ell_{im}^2$ = communality

ψ_i = uniqueness

The communality and uniqueness both constitute the amount of explained variance in a factor model. While communality is the portion of variance explained by the m-factors, uniqueness explains the portion of variance solely attributed to the variables $\{x_1, x_2, ..., x_p\}$; it gives the portion of variance not explained by the m-factors. This implies that: $\text{Uniqueness} = 1 - \text{Communality}$, [16].

RESULTS AND DISCUSSION

Eighteen events on the highway that could result to RTA as collated by FRSC include: Speed Violation (SPV), Use of Phone While Driving (UPWD), Tyre Burst (TBT), Mechanically Deficient Vehicle (MDV), Brake Failure (BFL), Overloading (OVL), Dangerous Overtaking (DOT), Wrongful Overtaking (WOT), Dangerous Driving (DGD), Bad Road (BRD), Route Violation (RTV), Road Obstruction Violation (OBS), Sleeping on Steering (SOS), Driving Under Alcohol/Drug Influence (DAD), Poor Weather (PWR), Fatigue (FTQ), Sign Light Violation (SLV) and Others.

Hence $p = 18$ and $n = 36$ (states of the federation including the FCT).

The random vector X in this study is thus:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_{18} \end{bmatrix}$$

Table 1: Random vector in the study

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}
Spv	Upwd	Tbt	Mdv	Bfl	Ovl	Dot	Wot	Dgd	Brd	Rtv	Obs	Sos	Dad	Pwr	Ftq	Slv	Others

Source: www.nigerianstat.gov.ng.

Table 2: Covariance Matrix

	Spv	Upwd	Tbt	Mdv	Bfl	Ovl	Dot	Wot	Dgd	Brd	Rtv	Obs	Sos	Dad	Pwr	Ftq	Slv	Others
Spv	4117.4																	
Upwd	9.7	0.40																
Tbt	245.6	0.6	32.1															
Mdv	132.4	0.3	9.0	16.4														
Bfl	94.2	0.4	7.0	9.3	33.8													
Ovl	42.7	0.3	3.1	0.6	-0.1	3.3												
Dot	21.0	0.2	1.3	0.9	0.5	0.4	0.9											
Wot	236.4	0.5	16.5	9.4	7.1	2.4	1.4	44.9										
Dgd	267.1	1.3	19.1	10.9	6.1	4.1	2.5	20.7	67.2									
Brd	24.2	0.10	2.0	2.1	1.3	0.4	0.6	3.1	3.5	2.7								

Rtv	351.8	1.0	27.1	11.8	10.8	3.3	2.1	19.0	36.6	1.4	50.4							
Obs	48.7	0.2	4.0	1.2	3.6	1.7	0.4	4.4	3.3	0.6	4.3	12.0						
Sos	12.8	0.09	1.4	0.8	1.0	0.1	0.01	1.5	1.7	0.01	1.4	0.4	2.1					
Dad	9.4	0.003	0.3	0.8	1.0	-0.1	0.06	0.7	1.0	0.1	1.0	0.4	0.004	0.4				
Pwr	4.2	0.002	0.3	0.1	0.1	0.05	0.04	0.1	0.3	0.1	0.4	0.07	0.01	0.02	0.12			
Ftq	66.5	1.8		2.9	5.7	1.7	2.03	4.9	14.9	0.4	14.5	1.2	1.0	-0.03	-0.001	60.8		
Slv	728.7	4.1	7.8	20.9	21.4	8.5	8.5	20.3	110.9	2.8	102.6	11.8	3.2	2.9	1.4	69.4	539.7	
Others	278.9	3.2	47.5	19.7	17.6	7.5	4.0	3.5	45.8	2.3	50.0	7.1	1.4	1.5	-0.02	86.4	139.0	321.28

Source: Author's computation (2024) with R

Table 3: Correlation Matrix

	Spv	Upwd	Tbt	Mdv	Bfl	Ovl	Dot	Wot	Dgd	Brd	Rtv	Obs	Sos	Dad	Pwr	Ftq	Slv	Others
Spv	1.00																	
Upwd	0.24	1.00																
Tbt	0.68	0.17	1.00															
Mdv	0.51	0.12	0.39	1.00														
Bfl	0.25	0.11	0.21	0.39	1.00													
Ovl	0.36	0.30	0.30	0.08	-0.01	1.00												
Dot	0.34	0.25	0.25	0.22	0.08	0.24	1.00											
Wot	0.55	0.12	0.43	0.35	0.13	0.20	0.22	1.00										
Dgd	0.55	0.25	0.41	0.33	0.13	0.27	0.32	0.38	1.00									
Brd	0.23	0.09	0.21	0.32	0.26	0.14	0.40	0.28	0.26	1.00								
Rtv	0.77	0.25	0.67	0.41	0.18	0.25	0.31	0.40	0.63	0.12	1.00							
Obs	0.22	0.09	0.21	0.09	0.12	0.27	0.12	0.19	0.12	0.10	0.18	1.00						
Sos	0.14	0.10	0.17	0.13	0.27	0.02	0.01	0.15	0.14	0.002	0.14	0.08	1.00					
Dad	0.24	0.01	0.09	0.34	0.07	-0.07	0.10	0.18	0.19	0.11	0.24	0.17	0.01	1.00				
Pwr	0.19	0.01	0.15	0.1	0.12	0.08	0.12	0.04	0.12	0.17	0.17	0.06	0.02	0.11	1.00			

Ftq	0.13	0.37	0.17	0.09	0.13	0.12	0.27	0.09	0.23	0.03	0.26	0.04	0.09	-0.01	-0.0004	1.00		
Slv	0.49	0.28	0.36	0.22	0.16	0.20	0.38	0.13	0.58	0.07	0.62	0.15	0.09	0.15	0.17	0.38	1.00	
Others	0.24	0.28	0.25	0.27	0.17	0.23	0.23	0.03	0.31	0.08	0.39	0.11	0.05	0.14	-0.002	0.62	0.33	1.00

Source: Author's computation (2024) with R

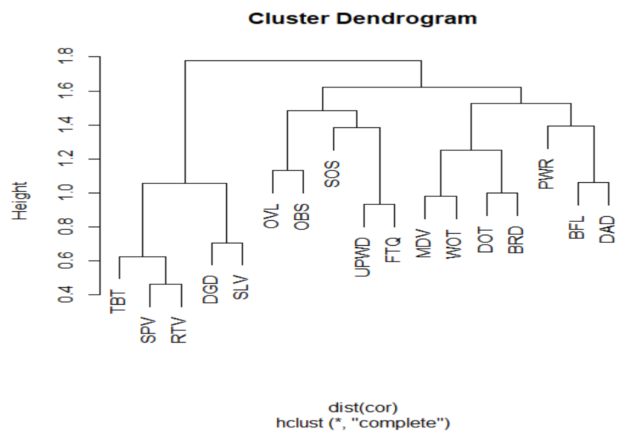

Fig 1: Dendrogram showing inter-relationship among the major causative factors of road traffic crash

Fig 1 summarizes the inter-relationship among some of the conventional causes of road accidents and it can be deduced that: rate at which SPV and RTV occur on Nigerian highway are similar and both of them can significantly result to TBT. The dendrogram also shows that motorists who engage in DGD are always committing SLV. Human factor which include UPWD and FTQ are both mapped to SOS (lack of concentration). A vehicle with a MDV status trying to engage in WOT is likely to experience road traffic crash. A motorist trying to do DOT on a BRD is prone to road accident. Most road traffic accidents caused as a result of BFL and DAD usually occur during PWR (e.g. wet season)

3.1 Principal component Analysis

3.1.1 Eigen decomposition of the correlation matrix

Table 4: Eigen values

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}	λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}
5.14	1.75	1.34	1.21	1.12	1.04	0.95	0.82	0.75	0.69	0.65	0.57	0.52	0.44	0.34	0.29	0.21	0.15

Source: Author's computation (2024) with R

Table 5: Eigen vector

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	-0.37	-0.17	-0.12	-0.18	-0.05	-0.005	-0.11	-0.12	0.09	0.006	0.03	-0.21	-0.02	0.04	0.12	-0.44	0.36	0.61
2	-0.18	0.35	-0.05	-0.10	0.26	0.08	0.10	0.01	0.75	-0.07	0.14	-0.03	-0.38	-0.05	-0.12	-0.03	-0.05	-0.04

3	-0.32	-0.12	-0.14	0.25	-0.004	0.06	-0.04	-0.31	-0.17	0.08	0.06	-0.20	-0.29	-0.43	0.008	0.52	-0.26	0.10
4	-0.260	-0.27	0.28	-0.13	0.04	0.20	-0.08	-0.22	0.05	-0.1	-0.33	-0.11	-0.21	0.64	0.12	0.23	0.01	-0.13
5	-0.16	-0.19	0.46	-0.09	0.26	-0.07	0.11	-0.26	0.19	0.54	-0.01	0.23	0.36	-0.18	-0.09	-0.008	-0.01	0.05
6	-0.18	-0.15	-0.46	-0.01	0.31	-0.21	-0.07	-0.17	-0.12	-0.20	--0.42	0.02	0.48	-0.04	0.22	0.10	0.01	-0.13
7	-0.23	0.12	-0.18	-0.45	-0.10	0.12	0.02	0.27	-0.09	0.31	0.02	-0.56	0.24	0.08	-0.31	0.11	-0.002	0.001
8	-0.24	-0.28	-0.16	0.07	0.19	0.28	-0.14	0.06	-0.005	-0.21	0.60	0.14	0.30	0.20	-0.01	-0.1	-0.33	-0.12
9	-0.32	0.04	-0.07	0.09	-0.22	0.06	-0.05	0.34	0.04	-0.08	-0.16	0.57	0.06	0.04	-0.39	0.33	0.18	0.23
10	-0.16	-0.19	-0.19	-0.60	0.04	0.29	0.10	0.03	-0.19	-0.007	-0.12	0.32	-0.26	-0.32	0.25	-0.22	0.04	0.07
11	-0.37	-0.02	0.019	0.27	-0.22	-0.07	-0.10	-0.03	-0.03	0.08	-0.002	-0.04	-0.07	-0.14	-0.16	-0.27	0.31	-0.71
12	-0.14	-0.08	-0.09	-0.03	0.53	-0.57	0.04	0.24	-0.31	0.15	0.13	0.07	-0.33	0.21	-0.08	0.02	0.07	-0.001
13	-0.09	-0.04	0.10	0.36	0.28	0.34	0.67	0.30	-0.13	-0.12	-0.22	-0.15	0.07	-0.06	0.02	-0.07	0.005	-0.01
14	-0.14	-0.28	0.42	-0.18	-0.03	-0.31	-0.13	0.38	0.16	-0.46	0.001	-0.23	0.10	-0.33	0.14	0.12	0.0003	-0.0007
15	-0.1	-0.14	-0.12	-0.19	-0.37	-0.37	0.65	-0.33	0.05	-0.21	0.19	0.03	0.06	0.12	-0.11	0.02	-0.03	0.0009
16	-0.18	0.51	0.20	-0.09	0.05	0.10	0.07	-0.10	-0.22	0.09	0.39	0.05	0.11	0.03	0.37	0.26	0.43	0.005
17	-0.29	0.21	0.03	0.11	-0.33	-0.18	0.06	0.29	0.05	0.31	-0.07	0.10	-0.02	0.14	0.51	-0.10	-0.47	0.01
18	-0.22	0.40	0.30	-0.09	0.07	-0.04	-0.10	-0.22	-0.33	-0.30	-0.16	0.02	-0.009	-0.04	-0.35	-0.34	-0.38	0.12

Source: Author's computation (2024) with R

From table 4, the first-six Eigen values are larger than 1, this is an indication that the variation in the dataset can adequately be attributed to the first-six Principal components

3.2 Principal Components

Table 6: Important Components

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18
Standard deviation	0.48	0.41	0.32	0.28	0.25	0.24	0.19	0.18	0.17	0.15	0.13	0.12	0.10	0.08	0.07	0.05	0.03	0.00

Proportion of Variance	0.27	0.19	0.11	0.08	0.07	0.07	0.04	0.04	0.03	0.03	0.02	0.02	0.01	0.01	0.01	0.00	0.00	0.00
Cumulative Proportion	0.27	0.46	0.57	0.65	0.72	0.79	0.83	0.87	0.90	0.93	0.95	0.97	0.98	0.99	1.00	1.00	1.00	1.00

Table 6 summarizes the percentage of variance in the dataset explained by each Principal component. It can be deduced that more than 70% of the variation in the data can be explained by the first-six Principal components (PC). Hence PC1 to PC6 are the important component in this study.

	PC1	PC2	PC3	PC4	PC5	PC6
SPV	**0.437**	0.170			0.102	
UPWD		-0.346		-0.209		**0.737**
TBT	0.391	0.121		0.182		0.303
MDV	0.201	0.267	-0.273	-0.162	-0.246	
BFL	0.191	-0.421		-0.178	-0.125	0.308
OVL	0.128	-0.153	**0.480**	0.174	-0.120	0.285
DOT	0.148	-0.121	0.223	-0.439	-0.115	
WOT	0.241	0.283	0.140	0.115	-0.310	-0.126
DGD	0.372			0.111		-0.165
BRD	0.194	0.270	**-0.499**	-0.286	-0.194	
RTV	**0.476**		0.106		0.168	
OBS		0.159	0.281		**0.602**	-0.179
SOS		**0.513**		**-0.570**	-0.226	-0.145
DAD	0.277	-0.382	-0.185		0.330	-0.396
PWR	0.141	0.158	-0.203	**0.681**		0.398
FTQ	**-0.507**	-0.226		-0.107		0.110
SLV	0.340	-0.217			0.342	
OTHERS	0.134	-0.401	-0.288		-0.125	0.133

Table 7: Loadings of Principal Component of Causative Factors on Road Accident in Nigeria

Table 7 contains the loadings of each variable in the six important PCs. According to [13], loadin gs define what principal components represent.

From table 7, the first has a high positive loading on SPV (0.437) and RTV (0.476) and a strong negative loading on FTQ (-0.507). SOS with a loading of 0.513 significantly affects PC2. OVL (0.480) and BRD (-0.499) are the important variables in PC3. PC4 is controlled by PWR (0.681) but also displayed a negative loading on SOS (-0.570), OBS (0.602) is of relative importance in PC5 and UPWD (0.737) is solely dominant in PC6.

FACTOR ANALYSIS

Table 8: Likelihood-ratio test for number of sufficient factors

Number of factor (p)	Test Statistic	Degree of freedom	p-value
1	991.69	135	2.05e-130
2	663.16	118	7.98e-77
3	498.97	102	1.3e-53
4	362.33	87	2.76e-35
5	255.44	73	4.49e-22
6	185.27	60	1.05e-14
7	116.87	48	1.13e-07
8	69.25	37	0.00103
9	38.98	27	0.0636

The likelihood ratio test for $p = 9$

The result from the table above implies that 9 is sufficient to describe the causes of road traffic accidents on Nigerian highway.

Table 9: Proportion of variance explained by the nine factors

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
SS loadings	1.372	1.370	1.274	1.219	1.185	1.148	1.049	1.035	0.931
Proportion of Variance	0.076	0.076	0.071	0.068	0.066	0.064	0.058	0.058	0.052
Cumulative Proportion	0.076	0.152	0.223	0.291	0.357	0.420	0.479	0.536	0.588

Table 10: Factor loadings

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	**Uniqueness**

SPV		**0.403**	0.255	**0.763**	0.302		0.115	0.244		0.007
UPWD	0.412	0.105		0.100			0.260			0.727
TBT		0.256	0.111	0.272	**0.842**		0.130	0.195		0.061
MDV		0.528	0.286	0.193	0.223		0.150			0.505
BFL	0.116		**0.614**		0.127					0.583
OVL	0.113		-0.114	0.158		0.110	`			0.265
DOT	0.267	0.316		0.120		0.389		0.156		0.628
WOT	0.108		0.228	**0.466**	0.281	0.212	0.206	0.127	-0.145	0.514
DGD	0.153	0.304	0.132	0.186	0.164	0.152	**0.868**	0.116	0.100	0.005
BRD			0.128			**0.897**				0.161
RTV	0.137	**0.496**	0.247	0.396	0.397		0.292	0.118	0.217	0.214
OBS			0.213					0.353		0.811
SOS	0.113		0.133		0.163					0.926
DAD		0.103	**0.506**							0.703
PWR		0.288				0.152				0.909
FTQ	**0.841**								0.252	0.210
SLV	0.328	**0.761**	0.119				0.278			0.183
OTHERS	0.460		0.177					0.131	**0.847**	0.005

Table 9 contains the loadings of the nine sufficient factors and the uniqueness of the 18 variables in this study in the nine factors. It can be observed that FTQ (0.841) has a large positive loading on factor 1, so this factor describes motorist level of exhaustion at a point in time. SPV (0.403), RTV (0.496) and SLV (0.761) have positive loadings on factor 2, this implies that factor 2 can be viewed as a measure of road traffic rules violation. BFL (0.614) and DAD (0.506) have positive loadings on the third factor, hence factor 3 is a measure of loss vehicular control. SPV (0.763) and WOT (0.466) have high positive loadings on factor 4, this indicates that factor 4 describes driver's recklessness acts on the high way. The structural integrity of tyres is measured by Factor 5. BRD (0.897) has a large positive loading on factor 6, factor 6 describes bad road conditions in Nigeria. OVL (0.798) and DGD (0.868) have high loadings on factor 7, hence factor 7 describes the profit-making intentions of commercial motorist on the highway. OTHERS (0.847) is dominant in factor 9, factor 9 describes other causes of road accidents not captured by variables in this study (which may include teenage driving. Indiscriminate parking and pressure from passengers).

Nine, out of the 18 variables each have uniqueness below 50%. The variables are: SPV, TBT, OVL, DGD, BRD, RTV, FTQ, SLV, OTHERS. This implies that these variables are well represented by the 9 factors identified in this study.

Discussion and Conclusion

This paper revealed the major causes of road traffic accident on the Nigerian Highway using multivariate analytic approach which include the principle component analysis for data reduction and the factor analytic approach for thoughtful interpretation.

The correlation matrix was used in carrying out a principal component analysis on eighteen variables. The PCA revealed that the first six components crucially account for more than 70% of variation in the original eighteen variables. Loadings of each variable in the six PCs were inspected in order to identify the variables that are of relevance important. It was observed that the first PC exhibits a positive relationship with speed violation and Route violation, it also displayed a negative correlated with Fatigue (FTQ). This implies that the first PC will increase when: speed violation increases, route violation increases and fatigue decreases. This component can be said to describe the recklessness of Motorist. A positive relationship between the second PC and sleeping on steering (SOS) indicates that the second PC increases with an increase in SOS. The second PC can be viewed as a measure of erratic driving. The third PC showed a positive correlation with overloading (OVL**)** and a negative relationship with bad road (BRD), this indicates that when overloaded vehicles ply a road too often, such road is prone to damage. The third PC can be viewed as a measure of profit made by commercial motorists and haulage companies (The more passengers/goods on-board, the more profit they are likely to make). The fourth PC showed a strong positive relationship with poor weather condition (PWR**)** and negative relationship with sleeping on steering (SOS), this implies that the PC increases with an increase in poor weather condition and a decrease in rate at which drivers sleep on steering during poor weather. The PC is a measure of mental alertness of motorist during poor weather. The fifth PC obviously measures unanticipated road obstructions on the highway. The sixth PC increases with an increase in Use of Phone while driving (UPWD), the PC is a measure of motorist's cognitive distraction on the highway.

An exploratory factor analysis was also carried out to buttress the results from the PCA. A likelihood-ratio test was carried out to determine the number of sufficient factors that can model the road traffic crash data. The result from the test showed that no fewer than nine (9) factors can adequately model the causes of road accidents on the Nigerian Highway. The amount of variance in the dataset explained by the nine factors was examined and it was observed that more than 50% of the variance is explained by the nine factors. Loadings of all the variables in the nine factors were obtained and summarize in table 10. The result from the factor analysis showed that: exhaustion level of motorists, traffic rules violation, greed of commercial motorist to maximize profit, structural integrity of tyres, bad road conditions, recklessness of motorists on the highway and others (which may include teenage driving. Indiscriminate parking and pressure from passengers) constitute the factors contributing to road traffic accidents on Nigerian highways. The identified factors can adequately be quantified by: speed violation (SPV), Tyre burst (TBT), Overloading (OVL), Dangerous driving (DGD), Bad road (BRD), Route violation (RTV), Fatigue (FTQ), Sign Light violation (SLV) and OTHERS.

Based on the findings in this study, we recommend that the road traffic agencies should focus their interventions on the top contributing factors by implementing stricter traffic regulations to curb motorist reckless behaviours on the highway through adoption of more sophisticated technology. There is need for the government at various tiers to improve their road maintenance culture, so as to reduce incidents related to mechanical failures and tyre bursts. Frequent extension programmes should be organized for commercial and haulage drivers to emphasize the dangers of overloading and high risk-driving behaviours.

References

[1] Akinyemi, E. O. (2009). Contributing Road Factors in Accidents on Rural Roads in Nigeria: In Road Traffic Accidents in Developing Countries", *Joja press*, Lagos.

[2] Andrilli, S. and Hecker, D. (2010): "Orthogonality". Elementary Linear Algebra, *Elsevier:* pp.397-44

[3] Atubi, A.O. (2010). Road Transport System Management and Traffic in Lagos, South Western Nigeria. *Journal of African Research Review*, 4(4):459-470.

[4] Aworemi, A. and Alawode, E.C. (2009). Roads to where? *News watch*, 17 (13): 3.

[5] Christof, S. and Ke-Hai, Y. (2005): Factor Analysis. *Encyclopedia of Social Measurement*, 2 ©2005

[6] Greenacre, M., Groenen, P.J.F., Hastle, T., Denza,A.I., Markos A. and Tunzhilina E.(2022): Principal component analysis. *Nature Reviews methods primer*, (2022) 2:100.

[7] Gungul, T. (2012). Road Traffic Accidents in Nigeria: Causes and Control. *Journal of Applied Science and the Environment*, 3: 25-29.

[8] Haddon, W. (1980). Advances in the Epidemiology of Injuries as a basis for public policy. *Public Health Reports* 95: 411-421.

[9] Krug, E.G., Sharma, G.K.; Lozano, R. (2000). The Global Burden of Injuries. *American Journal of Public Health.* 90(4): 523-526.

[10] National Beaureu of Statistics [NBS] (2020). Road Transport Data. *NBS report*, September, 2020.

[11] Olagunju, K. (2011) Road sense, *verigab Nig.* Ltd Abuja.

[12] Pishro-Nik, H., (2014). Introduction to probability, statistics and random processes. *Kappa Research LLC,* 2014.

[13] Rasmus B.and Age K.(2014): Principal Component Analysis. *Anal. Methods*, 2014, 6: 2812-2831

[14] Richard, A. Johnson and Dean W. Wichern (2007) : "Applied Multivariate Statistical Analysis. Pearson Education, Inc. New Jersey.

[15] Sanjay, K.S. (2017). "Road Traffic Accidents in India: Issues and Challenges". *Transportation Research Procedia* 25 (2017): 4708–4719.

[16] Sarmentos, R. and Costa, V. (2017): FACTOR ANALYSIS. In book: Comparative Approaches to Using R and Python for Statistical Data Analysis. DOI: 10.4018/978-1-68318-016-6.ch007. pp. 116-140.

[17] Sirwan K. A., Mona G. M., Salar O. A., Rabab G. A., Nahed A. E., Deepak C., Mohammad E. U., Kuldeep D.(2023). Road traffic accident injuries and deaths: A neglected global health issue. *Health Science Report*, 6 (5):1-6.

[18] Transport and Research Laboratory [TRL] (1999). "Towards Safer Roads in Developing Countries", *Transport and Research Laboratory (TRL)*, Crowthorne, U.K.

[19] Wasiu, A.Y., Mutiu, S., Barau, R. "Road Traffic Accidents in Nigeria-Trends, Challenges and Forecast" Conference paper: International Conference for Social Sciences, Nile University of Nigeria, FCT, Abuja.

[20] Weisstein,E.W.(2024)."Normalized vector". From math world – *A wolfram web resource*. https://mathworld.wolfram.com/Normalized vector.html.

[21] World Health Organization (WHO). "Global Status Report on Road Safety 2018". *Worl d Health Organization.* Geneva. 2018