STATISTICAL ANALYSIS OF CRIME DATA USING PRINCIPAL COMPONENT ANALYSIS: A CASE STUDY OF FCT ABUJA

Ibrahim Zakariyau[1*] and Yusuf Ibrahim[2]

[1]Tertiary Education Trust Fund (TETFund)

[2]Nigerian Defence Academy (NDA), Kaduna

[*]*Corresponding author Email: abuaskari@gmail.com*

ABSTRACT

Principal Component Analysis (PCA) is a statistical tool used for reducing the dimensionality of a very large data set and for determining the areas with the overall crime rate. The paper statistically analyzed the FCT Abuja crime data recorded and reported by the Nigeria Police Force using correlation analysis and principal component analysis for the period of 2018-2023. The crimes consist of robbery, theft/stealing, house breaking, store breaking, burglary, grievous hurt and wounding (G.H.W.), murder, rape, kidnapping, and assault. The result of the study has shown that theft has the highest rate and murder has the lowest rate. It was evidently shown that kidnapping has shown a significant correlation with grievous harm and wounding (GHW), which means that the high rate of GHW in FCT is associated with kidnapping and therefore each crime can be used to predict (explain) one other. PCA suggested retaining four components, namely, murder, burglary, assault, kidnapping, and armed robbery, that explain about 95% of the total variability of the crime data set. It was recommended among others that there is a need to tackle kidnapping to reduce crime, grievous harm and wounding (GHW), and routine training of Police personnel on advanced data collection, processing, analysis, and visualization.

Keywords: Analysis, Statistical Analysis, Crime, Correlation Analysis, Principal Component Analysis

1. INTRODUCTION

Crime is defined as the breach of criminal law that governs a particular geographical area, the criminal law that aims to protect the lives, property, and rights of citizens within a specific jurisdiction (Lamond, 2007). The degradation of Nigeria's economy is a great determinant in the high rate of crime occurrences (Danbazau, 2017). The application of multivariate statistics has made some contributions to many criminological explanations (Kpedekpo and Arya, 1981, and Printcom, 2003). This study explores the use of correlation analysis and principal component analysis (PCA) for effective crime control and prevention.

Correlation is the value of association between two independent or one independent and another dependent variable, determined by measuring the correlation coefficient (Pearson, Kendall, Spearman) and also the direction of their relationship, i.e., positive correlation or negative correlation. Quantification of this association involves computing the correlation coefficient ranging between $[-1,1]$ (Madhuri *et al*., 2023). Various types of crime can be analyzed using the correlation coefficient to determine the strength of relations between pairs of crimes.

Principal component analysis (PCA) is a novel technique for multivariate data analysis and the model prediction (Ringner, 2008). The goal of PCA is to reduce the number of variables by extracting several principal components. In general, the first components can capture most of the information in data. PCA has been proved to be an effective method for crime analysis (Yao, 2016).

PCA offers a tool for reducing the dimensionality of a very large data set and in determining the areas with overall crime rate. Correlation analysis and PCA are essential for effective crime control and prevention (Gumbe *et al*., 2013). Principal component analysis can also be used to determine the overall criminality. When the first eigenvector shows approximately equal loadings on all variables, then the first PC measures the overall crime rate (Faweya, 2018). Considering the progression of criminal activities in the Nation capital, this study aims to examine the distribution of various crime, the degree of relationship between the different crimes, determine the crime that accounted for highest percentage of the total crimes in FCT Abuja and to determine which of the acts constitute the major or principal components of the crimes as well as their effects on life in the study area.

Several studies have applied Principal Component Analysis (PCA) to crime data in various regions. Gulumbe (2013) and Atanu (2019) both found significant correlations between certain crimes, with Gulumbe identifying four components that explained 78.94% of the data variability and Atanu identifying two components that explained 85.166%. Erick (2019) focused on the causes of crimes in Mathare slums, Nairobi, and found a strong positive relationship between unemployment and drugs, with three components explaining 52.6% of the variability in crimes against persons and two components explaining 42.2% of the variability in crimes against property. Olakorede (2017) analyzed crime data in Gwagwalada Area Command, Abuja, and identified three components that explained 86.873% of the data variability. These studies collectively demonstrate the utility of PCA in identifying underlying patterns and relationships in crime data. Wafula (2019) conducted a statistical analysis of causes of crimes in Mathare slums, Nairobi County, using data collected via questionnaires in April 2018, using correlation analysis to explain the association between the causes of crimes and principal component analysis (PCA) to reduce the dimensionality of the data sets. Faweya *et al*. (2018), employed Principal Component Analysis (PCA) to reduce the dimensionality, and the results revealed that the first five principal variables which explained 66.769% of the total variation, and the first six principal variables which accounted 93.881% of the total variation in Ekiti State and Osun State data set, respectively have been retained using the scree plot. Olakorede *et al*. (2019) conducted a study that analyzed the crime data of Gwagwalada Area Command of FCT, consisting of the averages of twenty major crimes reported to the police for the period 1995 – 2015. The study also explored the use of correlation analysis and principal component analysis (PCA) to explain the correlation between the crimes and to determine the distribution of the crimes over the three Area Councils under the Gwagwalada Area Command. Atanu (2019) analysed the crime data of Nigeria, consisting of eight major crimes reported to the police in 2017. The study also employed the used Principal Component Analysis (PCA) to explain the correlation between the crimes. Musa *et al*. (2020) carried out a study using multivariate approaches of Principal Component Analysis and multidimensional scaling using secondary data collected from the 23 Divisional Police Headquarters (DPHs) on all the 23 Local Government Areas (LGA) of Kaduna State.

Contemporary literature showed there appears to be a gap in the patterns and relationships in crime data in certain places and certain periods of time.

However, there hasn't been a study that statistically analyzed the FCT Abuja crime data recorded and reported by the Nigeria Police Force using correlation analysis and principal component analysis for the period of 2018-2023. Therefore, the need to examine crime data of FCT Abuja is highly appropriate. This paper statistically analysed the FCT Abuja crime data recorded and reported by the Nigeria Police Force using correlation analysis and principal component analysis for the period of 2018-2023.

2. LITERATURE REVIEW

Many studies have been conducted on the use of principal component analysis to analyze crime data in various places for certain periods of time. However, in this paper, contemporary literatures on statistical analysis of crime data using principal component analysis were reviewed.

In Kenya, Wafula (2019) conducted a statistical analysis of causes of crimes in Mathare slums, Nairobi County, using data collected via questionnaires in April 2018, using correlation analysis to explain the association between the causes of crimes and principal component analysis (PCA) to reduce the dimensionality of the data sets. The result of the study showed that there was a fairly strong positive relationship between unemployment and drug and substance abuse, which means that their variables can be used to predict one another. It also revealed that three PCs (drugs and substance abuse, unemployment, and neglect from parents) that explain about 52.6% of the total variability of the causes of crimes against persons are suggested to be retained. Similarly, two PCs (drugs and substance abuse and unemployment) that explain about 42.2% of the total variability of the causes of crimes against property are suggested to be retained.

Musa *et al*. (2020) carried out a study using multivariate approaches of Principal Component Analysis and multidimensional scaling using secondary data collected from the 23 Divisional Police Headquarters (DPHs) on all the 23 Local Government Areas (LGA) of Kaduna State. Findings of the study revealed that three components were retained, which collectively accounted for about 85% of the total variability in the data, and individually, the first, second, and third PC accounted for 42.4%, 25.5% and 16.9% of the total variability in the data, respectively. It further revealed that the Euclidean Distance map employed in the study shows that grievous hurts, assaults, rape, murder, and kidnapping are closely compact.

A study conducted by Atanu (2019) analysed the crime data of Nigeria, consisting of eight major crimes reported to the police in 2017. The study also employed the used Principal Component Analysis (PCA) to explain the correlation between the crimes. The result of the study showed a significant correlation between armed robbery, theft and grievous harm, and wounding. It further showed that Lagos has the highest crime rate, and Kebbi state has the lowest. Also, theft, house breakings, grievous hurt and wounding, murder, rape are more prevalent in Lagos state, armed robbery in Rivers State, while kidnapping and assault are prevalent in Abia state. The PCA has suggested retaining two components that explain about 85.166 percent of the total variability of the data set.

Gulumbe *et al*. (2013) analysed crime data of Katsina State consisting of eight major crimes reported to the police for the period 2006 – 2008. The study employed the use of correlation analysis and principal component analysis (PCA) to explain the correlation between the crimes and to determine the distribution of the crimes over the local government areas of the state. Findings of the study revealed that there was a significant correlation between robbery, theft, and vehicle theft. Katsina local government recorded the highest crime rate, and Musawa local government recorded the lowest. The findings further showed that robbery is more prevalent in Danmusa local government area, rape in Jibia local government area, and grievous hurt and wounding in Dandume local government area. The PCA used in the study suggested retaining four components that explain about 78.94 percent of the total variability of the data set.

In Sokoto, Usman *et al*. (2012) applied the PCA technique to determine the number of principal components to be retained on the seven variables obtained from the Criminal Investigation Department, Sokoto State Police Headquarters, Sokoto. Findings of the study showed that three Principal components that explain about 89 percent of the total variability of the data set have been retained using the Scree plot and Loading plot, indicating that correlation exists between crimes against persons and crimes against property.

In a study conducted by Faweya *et al*. (2018), the statistical analysis revealed that the commonly committed crimes out of others in Ekiti State are theft and stealing, store and office breaking, grievous harm, house breaking and assault occult harm and the highest and commonly committed crimes in Osun State are; house breaking, receiving stolen property, armed robbery, assault, grievous harm and wounding, arson. The Principal Component Analysis (PCA) employed in the study was used to reduce the dimensionality, and the results revealed that the first five principal variables, which explained 66.769% of the total variation, and the first six principal variables which accounted 93.881% of the total variation in Ekiti State and Osun State data set, respectively have been retained using the scree plot.

Olakorede *et al*. (2019) conducted a study that analyzed the crime data of Gwagwalada Area Command of FCT, consisting of the averages of twenty major crimes reported to the police for the period 1995 – 2015. The study also explored the use of correlation analysis and principal component analysis (PCA) to explain the correlation between the crimes and to determine the distribution of the crimes over the three Area Councils under the Gwagwalada Area Command. The result of the study showed a significant correlation between robbery and rape, grievous hurt and wound (GHW), theft, assault, murder, and unlawful escape. It also revealed that the highest overall crime rates are Rape, Robbery, GHW, Theft, Assault, Murder, and unlawful escape in the Area Command. Also, unlawful possession, breach of peace, and broken store are more prevalent in the Kwali Area Council, while Vehicle theft, car stealing and burglary are more prevalent in the Kuje Area Council Area. Findings of the study further revealed that the PCA suggested retaining three components (Rape, Robbery and GHW) that explain about 86.873 percent of the total variability of the data set.

Several studies have applied Principal Component Analysis (PCA) to crime data in various regions. Gulumbe (2013) and Atanu (2019) both found significant correlations between certain crimes, with Gulumbe identifying four components that explained 78.94% of the data variability and Atanu identifying two components that explained 85.166%. Erick (2019) focused on the causes of crimes in Mathare slums, Nairobi, and found a strong positive relationship between unemployment and drugs, with three components explaining 52.6% of the variability in crimes against persons and two components explaining 42.2% of the variability in crimes against property. Olakorede (2017) analyzed crime data in Gwagwalada Area Command, Abuja, and identified three components that explained 86.873% of the data variability. These studies collectively demonstrate the utility of PCA in identifying underlying patterns and relationships in crime data.

However, there are no published data on statistically analyses of FCT Abuja crime data recorded and reported by the Nigeria Police Force using correlation analysis and principal component analysis for the period of 2018-2023.

3. METHODOLOGY

The study is FCT Abuja and the data used in this study was collected from the Research and Planning Department, FCT Abuja Police Command. The data consists of ten selected major crimes reported to the police for the period 2018 – 2023. The crime classifications are: Crimes against properties which include robbery, theft/stealing, house breaking, store breakings and burglary, and crimes against persons which include grievous hurt and wounding (G.H.W.), murder, rape, kidnapping and assault.
The crime data obtained from the Research and Planning Department, FCT Abuja Police Command contained 24 crime types, namely: murder, arm robbery, kidnapping, rape, false pretense & cheating, one chance, abduction, forgery, assault, grievous harms and wounding, public nuisance and child stealing. Others are theft & stealing, house breaking, store breaking, burglary, criminal misappropriation, indecent assault, criminal breach of trust & cheating, criminal intimidation, disturbance of public peace, homicide, receiving stolen property and defamation of character. For each crime type there are 8 variables, namely: cases reported, number of persons arrested, number of persons charge to court, number of persons convicted, number of persons awaiting trial, number of arms & ammunition recovered, exhibit recovered and location of crime.

The crime type was cleaned by trimming the number of crime type to 10 such as robbery, theft/stealing, house breaking, store breakings, burglary, grievous hurt and wounding (G.H.W.), murder, rape, kidnapping and assault. Also, for the variable, only the number of cases reported was used.
PCA is employed to determine the distribution of the crimes over the Area Councils of FCT Abuja. The correlations among different crimes are explained based on correlation analysis.

The data consists of ten major crimes reported to the police for the period 2018 – 2023 namely, robbery, theft/stealing, house breaking, store breaking, burglary, grievous hurt and wounding (G.H.W.), murder, rape, kidnapping and assault. Frequencies of crimes for each category were averaged over the three years in the study period to control for anomalous years when there may have been an unexplained spike or fall in crime levels prior to the statistical analysis.

Thus, the crime data set contained the 10 crime types of the crime and the number of cases reported for each year. The objective of this study includes: reduce the dimensionality of the crime data by using PCA; identify the major crime types in FCT; determine the association existing between different crime types. The analysis procedure is organized as follows: First, the eigenvalues for different components are calculated. Then, the principal components for raw data are computed. Finally, the scores of each observation for the pair principal components are plotted.

The correlation or correlation coefficient (r) is a pure number between -1 and 1 summarizing the strength of the relationship. A correlation of 1 indicates a perfect straight-line relationship with upward tilt; a correlation of -1 indicates a perfect straight-line relationship with downward (negative) tilt. The correlation tells you how close the points are to being exactly on a tilted straight line, but it does not tell you how steep that line is. The formula for the correlation coefficient is:

$$r = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y})}{S_X S_Y} = \frac{covariance}{S_X S_Y}$$

The **covariance** of *X* and *Y* is the numerator in the formula for the correlation coefficient. Because its measurement units are difficult to interpret, it is probably easier to work with the correlation coefficient instead (Siegel, 2022).

Principal component analysis was be used as a statistical procedure to transform a set of observations of correlated variables to a set of a linearity uncorrelated variables by orthogonal transformation. It can be explained that the variance covariance structure of these variables by some of these linear combinations of the original variables.

PCA calculates an uncorrelated set of variables (principal components), These factors are ordered so that the first few retain most of the variation present in all of the original variables. Unlike its cousin Factor Analysis, PCA always yields the same solution from the same data (apart from arbitrary differences in the sign).

Let *X* be a vector of *p* random variables, the main idea of the PC transformation is to look for a few *(<p)* derived variables that preserved most of the information given by the variance of the *p* random variables (Jolliffe, 2002).

Let the random vector $X^{/} = [X_1, X_2, \ldots, X_p]$ have the covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq, \cdots \geq \lambda_p \geq 0$.

Consider the linear combinations,

$Y_j = \alpha'_{j1}X_1 + \alpha'_{j2}X_2 + \cdots + \alpha'_{jp}X_p = \alpha \sum \alpha'_{jk}X_k;\ j = 1,2,\cdots,p$ of the element of X, where $\propto_j$ is a vector of p components $\alpha'_{j1}, \alpha'_{j2}, \cdots, \alpha'_{jk}$.

Then,

$$Var(Y_j) = \alpha'_j \Sigma \alpha_j \qquad j = 1,2,\cdots,p,\cdots$$

$$Var(Y_j) = \alpha'_k \Sigma \alpha_j \qquad k = 1,2,\cdots,p,\cdots$$

The PC's are those uncorrelated linear $Y_1, Y_2, \cdots, Y_p$ whose variances are as large as possible (Richard and Dean, 2002). In finding the PC's we concentrate on the variances. The first step is to look for a linear combination $\alpha'_j X$ with maximum variance, so that

$$\alpha'_{jX} = \alpha_{11}X_1 + \alpha_{12}X_2 + \cdots + \alpha'_{1p}X_p.$$

Next, look for a linear combination $\propto'_2 X$ uncorrelated with $\alpha'_j X$ having maximum variance, and so on, so that at the K^{th} stage a linear combination $\alpha'_k X$ is found that has maximum variance subject to being uncorrelated with $\alpha_1'X + \alpha_2'X + \cdots + \alpha_{k-1}'X$. The kth derived variable $\alpha_k'X$ is the K^{th} PC. Up to p PC's could be found, but we have to stop after the q^{th} stage $(q \leq p)$ when most of the variation in X have been accounted for by q PCs.

The variance of a PC is equal to the eigenvalue corresponding to that PC,

$$Var(Y_j) = \alpha_j'\Sigma\alpha_j \qquad\qquad j = 1,2,\cdots,P$$

The total variance in a data set is equal to the total variance of PCs
$\sigma_{11} + \sigma_{12} + \cdots + \sigma_{pp} = \Sigma Var(Y_j)$

The data was standardized for the variables to be similar scale using a common standardization method of transforming all the data to have zero mean and unit standard deviation. For a random vector $X' = [X_1, X_2, \cdots, X_p]$ the corresponding standardized variables are

$$Z = \frac{X_j - \mu_j}{\sqrt{\sigma_{11}}} \qquad\qquad j = 1,2,\cdots,p$$

In matrix notation
$Z = V^{\frac{1}{2}}(X - \mu)$

where $V^{1/2}$ is the diagonal standard deviation matrix.

Thus, $E(Z) = 0$ and $Cov(Z) = P$. The PCs of Z can be obtained from eigenvectors of the correlation matrix ρ of X. All over previous properties for X are applied for the Z, so that the notation Y_j refers to the j^{th} PC and (λ_j, α_j) refers to the eigenvalue- eigenvector pair. However, the quantities derived from Σ are not the same as those derived from ρ (Richard and Dean, 2001 [14]).
The j^{th} PC of the standardized variables $Z = [Z_1, Z_2, \cdots, Z_p]$ with $Cov(Z)$, is given by

$$Y = \alpha_j'Z = \alpha_j'\left(V^{\frac{1}{2}}\right)(X - \mu).$$

so that,

$$\sum Var(Y_j) = \sum Var(Z_j) = P; \qquad\qquad j = 1,2,\cdots,p$$

In this case, $(\lambda_1, \alpha_1), (\lambda_2, \alpha_2), \cdots, (\lambda_p, \alpha_p)$ are the eigenvalue–eigenvector pairs for p with $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq, \ldots, \lambda_p \geq 0$.

Interpretation of Principal Component Model

The loading or the eigenvector $\alpha_j = \alpha_1, \alpha_2, \alpha_k$, is the importance of a measured variable for a given measure of the PC. When all elements are positive, the first component is a weighted average of the variables and is sometimes referred to as a measure of the overall crime rate. Likewise, the positive and negative coefficients in subsequent components may be regarded as a type of crime component (Rencher, 2002, and Louis, 1981). The plot of the first two and three loadings against each other enhances visual interpretation (Soren, 2006).

The score is a measure of the importance of a PC for an observation. The new PC observation Y_{ij} are obtained simply by substituting the original variables X_{ij} into the set of the first q PCs. This gives

$$Y_{ij} = \alpha_j' X_{j1} + \alpha_{j2}' X_{j2} + \cdots + \alpha_{jp}' X_{jp}; \qquad j = 1,2,\cdots,p; \qquad i = 1,2,\cdots,n$$

The plot of the first two or three PCs against each other enhances visual interpretation (Soren, 2006). The proportion of variance tells us the PC that best explained the original variables. A cumulative proportion of explained variance is a useful criterion for determining the number of components to be retained in the analysis. A scree plot provides a good representation of the ability of the PCs to explain in the data (Cattell, 1996).

3.7.2 The Proportion of Variance

The proportion of variance tells us the PC that best explained the original variables. A measure of how well the first q PCs of Z explain the variation is given by:

$$\varphi_q = \frac{\sum_{j=1}^{q} \lambda_j}{P} = \frac{\sum_{j=1}^{q} Var(Z_j)}{P}$$

A cumulative proportion of explained variance is a useful criterion for determining the number of components to be retained in the analysis. A Scree plot provides a good graphical representation of the ability of the PCs to explain the variation in the data (Cattell, 1966).

4. RESULTS AND DISCUSSION

The FCT crime data for the period of 2018 - 2023 were processed and analyzed using R software. Below are the results of the computations.

Table 1: Descriptive Statistics

Variable	Count	Mean	Standard Deviation
Murder	6	31.33	11.47
Arm Robbery	6	73.83	19.97
Kidnapping	6	55.33	20.27
Rape	6	41.67	7.45
Assault	6	138.67	42.41
GHW	6	103.50	21.92
Theft	6	2104.17	528.64
House Breaking	6	63.00	8.00
Store Breaking	6	53.67	29.71
Burglary	6	48.00	24.06

Table 2: Correlation Matrix

Variable	Murder	Arm Robbery	Kidnap ping	Rape	Assault	GHW	Theft	House Breaking	Store Breaking	Burglary
Murder	1.0000									
Arm Robbery	0.0378	1.0000								
Kidnapping	-0.1709	0.5632	1.0000							
Rape	0.6456	-0.1174	-0.7038	1.0000						
Assault	0.2824	0.2188	-0.3523	0.2972	1.0000					
GHW	0.3318	0.0039	0.5116	0.0037	-0.6137	1.0000				
Theft	-0.4538	0.3608	0.2989	-0.5795	-0.1156	-0.4726	1.0000			
House Breaking	-0.7958	0.0238	0.0234	-0.6344	0.1232	-0.7412	0.7430	1.0000		
Store Breaking	0.1436	0.1317	-0.1392	-0.1425	0.4936	-0.6746	0.6468	0.4771	1.0000	
Burglary	-0.0326	-0.2851	0.2091	-0.5045	-0.3862	-0.0144	0.5910	0.3024	0.5414	1.0000

Table 3: Eigenvalue of the Correlation Matrix

Component	Eigenvalue	Percentage of Variance	Cumulative Percentage
1	3.79805	37.98	37.98
2	2.66406	26.64	64.62
3	1.59727	15.97	80.59
4	1.44106	14.41	95.00
5	0.49956	5.00	100.00
6	0.00000	0.00	100.00
7	0.00000	0.00	100.00
8	0.00000	0.00	100.00
9	0.00000	0.00	100.00
10	0.00000	0.00	100.00

Table 4: Rotated Component Matrix

Crime	Component				
	1	**2**	**3**	**4**	**5**
Murder	0.96	0.11	0.19	-0.16	
Arm Robbery			0.14	0.20	0.97
Kidnapping			-0.27	0.85	0.44
Rape	0.55	-0.29	0.10	-0.78	
Assault			0.97	-0.16	0.11
GHW	0.57	-0.29	-0.64	0.43	
Theft	-0.55	0.73			0.40
House Breaking	-0.90	0.35	0.25		
Store Breaking		0.81	0.56		0.10
Burglary		0.89	-0.24	0.26	-0.28

Table 5: Eigenvectors

Crime	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Murder	-0.64	0.24	0.30	0.66						
Arm Robbery	0.18	-0.12	0.94	-0.14	0.25					
Kidnapping	0.28	-0.79	0.51		-0.20					
Rape	-0.76	0.54			0.36					
Assault		0.81	0.42	-0.10	-0.39					
GHW	-0.57	-0.79		0.20						
Theft	0.92		0.15	0.17	0.33					
House Breaking	0.92	0.18	-0.16	-0.31						
Store Breaking	0.61	0.56	0.19	0.52						
Burglary	0.56	-0.23	-0.32	0.73						

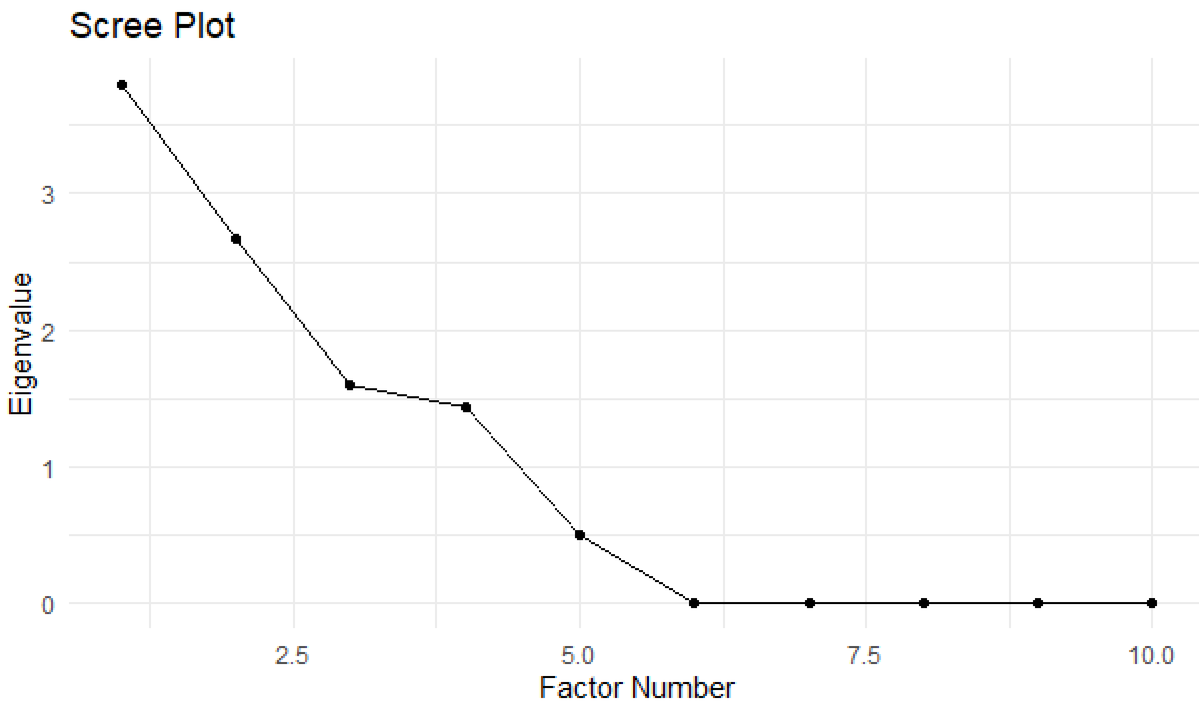

Figure 1: Scree Plot for FCT Crime Rate

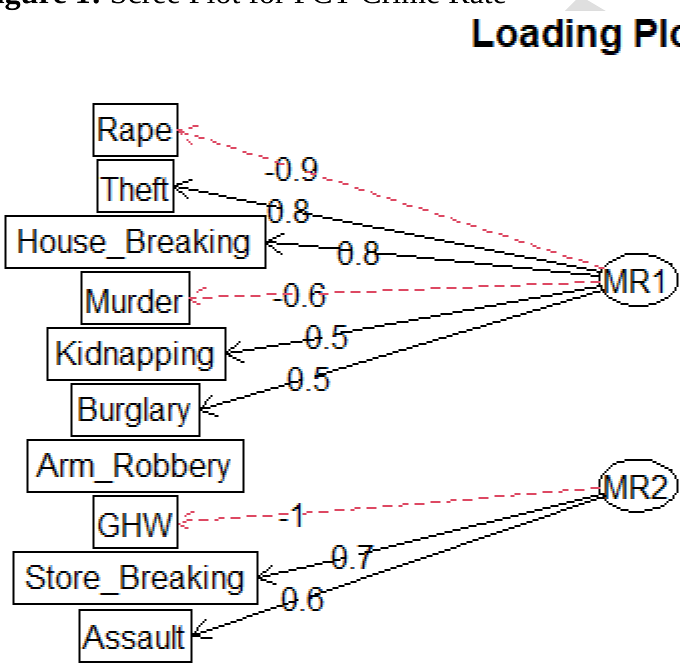

Figure 2: Loading Plot

DISCUSSIONS

Table 1 shows the descriptive statistics of the FCT Crime Data comprising of frequency, mean and standard deviation. In comparison of the relative sizes of the mean in the data set, theft has highest rate and murder has lowest rate.

Table 2, the correlation matrix has displayed different levels of (pairwise) correlation between different crime data variables. High correlations, usually greater than 0.5, indicate that the crimes may be related or influenced by similar factors. Looking at the table, rape has shown a very low correlations between kidnapping, theft, house breaking and burglary which means that none of the variable can be used to predict (explain) one another. It also means that the high rate of rape in FCT is not associated with property crime. However, kidnapping has shown a significant correlation with grievous harm and wounding (GHW), meaning that with kidnapping, grievous harm and wounding (GHW) can be predicted. Theft has shown a significant correlation with house breaking, store breaking, and burglary. This means the high rate of theft in FCT is associated with property crime and therefore each crime can be used to predict (explain) the others.

Eigenvalues are often used to determine how many factors to retain. Usually, factors with values greater than 1 are retained. From Table 3, the eigenvalues represent the amount of variance explained by each principal component. The first eigenvalue (3.79) explains 37.9% of the variance, the second eigenvalue (26.64) explains 26.64% of the variance, the third eigenvalue (1.6) explains 15.97% of the variance, the fourth eigenvalue (1.44) explains 14.41% of the variance, and the fifth eigenvalue (0.5) explains 5% of the variance. The cumulative proportion column showed the proportion of variance explained by each component and all previous components. Considering the eigenvalues and cumulative percent of the table, it will be reasonable to retain the first four PC's that explain up to 95.0% of the total variability in the data set.

Table 4, the rotated component matrix shows the loadings of each variable on the extracted factors. It contains columns representing the components, which are the underlying factors or dimensions that explain the correlations among variables, and rows representing the variables (crime type), which are the original measures or indicators. The values in the matrix represent the strength and direction of the relationship between each variable and each component. A high loading (close to 1) indicates that the variable is strongly related to the component, and a low loading (close to 0) indicates that the variable is weakly related to the component. The table showed that five components in each column have the highest value close to 1, namely: murder, burglary, assault, kidnapping, and armed robbery.

From Table 5, the eigenvectors represent the weights of the original variables in the factors. It also represents the direction of the new axes (principal components) in the transformed space. Each column of the eigenvectors table represents a principal component, and each row represents a variable. The values in the table represent the loadings of each variable onto each principal component. In the table, the first eigenvector (PC 1) has high loadings for theft (0.92), house breaking (0.92), store breaking (0.61), and burglary (0.56), indicating that theft, house breaking, store breaking, and burglary are strongly correlated with the first principal component. The second eigenvector (PC 2) has high loadings for rape (0.54), assault (0.81), and store breaking (0.56), indicating that rape, assault, and store breaking are strongly correlated with the second principal component. The third eigenvector (PC 3) has high loadings for armed robbery (0.94) and kidnapping (0.51), indicating that rape, assault, and store breaking are strongly correlated with the third principal component. The fourth eigenvector (PC 4) has high loadings for murder (0.66), store breaking (0.52), and burglary (0.72), indicating that rape, assault, and store breaking are strongly correlated with the fourth principal component

The scree plot helps in determining the number of factors to retain. The point where the plot levels off (the "elbow") suggests the optimal or important number of factors/components. Components after this contribute little to explaining variance. Figure 1 shows that the first four factors are indeed the largest, and we have the impression that these factors will adequately approximate this data and account for over 95% of the variation.

The loading plot visualizes the factor loadings and shows which variables strongly influence each other, thereby making it easier to interpret the relationships between variables and factors. Variables that are far from the origin have strong associations with components, and similar crimes may group to reveal patterns in the FCT crime data. Figure 2 classifies the crime into high and low crime rates, showing theft and house breaking as the highest crime rates, and murder with a low crime rate, then moderate positive loading on kidnapping and burglary.

5. CONCLUSION

In this study, correlation analysis and principal component analysis were applied to crime data. PCA was used to explore the number of principal components to be retained on the ten variables obtained from the Criminal Investigation Department of the FCT Police Command for the period of 2018 – 2023, which implies a great dimensionality reduction. The results of the statistical analysis proved that the four components, namely: murder, burglary, assault, kidnapping, and armed robbery, explain up to 95.0% (Table 4.3) of the total variability of the data set, thus suggesting that we do not lose much information.

6. RECOMMENDATION

Based on the findings of the study, the following were recommended:

i. Security operatives need to tackle kidnapping to avoid grievous hurt and wounding (G.H.W.)
ii. There is a need for a robust collection of crime data
iii. Police personnel need routine training on advanced data collection, processing, analysis, and visualization.

REFERENCE

Atanu, E. Y. (2019). Analysis of Nigeria's Crime Data: A Principal Component Approach using correlation matrix. International Journal of Scientific and Research Publications, 9(1), ISSN 2250-3153 http://dx.doi.org/10.29322/IJSRP.9.01.2019

Cattell, R.B (1996). The Scree test for the number of factors, Multivariate Behavioral Research, 1,245-276.

Danbazau, A.B (2007). Criminology Justice. 2nd edition. Ibadan: Spectrum Books Limited.

Faweya, O., Adeniran A. T. and Balogun K. O. (2018). Principal Component Analysis of Crime Rate in Nigeria: A Case Study of Ekiti and Osun State. American Journal of Mathematics and Statistics. 8(4): 79-88 doi:10.5923/j.ajms.20180804.01

Gulumbe, S. U., Dikko, H. G. and Bello, Y. (2012) : Analysis of crime data using principal component analysis: A case study of Katsina State, CBN Journal of Applied Statistics, ISSN 2476-8472, The Central Bank of Nigeria, Abuja, Vol. 03, Iss. 2, pp. 39-49

Jolliffe, I. T. (2002). Principal Component Analysis. 2nd edition, Springer-Verlag, New York. CBN Journal of Applied Statistics Vol. 3 No.2 49.

Kpedekpo, G.M.C. and Arya, P. L. (1981). Social and Economic Statistics for Africa. George Allen and Unwin, London.

Lamond, G. (2007). "What is a Crime? Oxford Journal of Legal Studies, vol. 27, no. 4, pp. 609-632, 2007.

Louis, S., Cookie, W.S., Louis, A. Z. and Sheldon, R. E. (1981). Human Response to Social problems. The Dorsey Press, Illinois.

Madhuri, G. and Atul, N. (2023). Statistical Modeling in Machine Learning. Discriminative dictionary learning based on statistical methods

Musa, G. A., Tasi'u, M., Alhaji, B. B. and Abdullahi, I. (2020). Determination of Crime Pattern in Kaduna State Using Principal Component Analysis and Multidimensional Scaling. Academy Journal of Science and Engineering (AJSE). Vol. 14 No.1. Retrieved from https://ajse.academyjsekad.edu.ng/index.php/new-ajse/article/view/50

Office of the Auditor General of the Federation [OAUGF] (2024). Abuja. Retrieved from https://www.oaugf.ng/6thawam2016/index.php/about/abuja#:~:text=Abuja%20is%20located%20in%20the,the%20country's%20most%20populous%20city.

Olakorede, N. M., Adams, S. O. and Olanrewaju, S. O. (2019). Principal Component Analysis of Crime Data in Gwagwalada Area Command, Abuja from 1995 – 2015. American Journal of Theoretical and Applied Statistics. Vol. 6, No. 1, 2017, pp. 38-43. doi: 10.11648/j.ajtas.20170601.15

Richard, A. J. and Dean, W. W. (2001). Applied Multivariate Statistical Analysis. 3rd edition, Prentice-hall, New Dehli.

Ringnér, M. (2008). What is Principal Component Analysis? Nature biotechnology, 26(3), 303.

Siegel, A. F. and Wagner, M. R. (2022). Correlation and Regression. Practical Business Statistics (Eighth Edition), 2022. Retrieved from https://www.sciencedirect.com/topics/mathematics/correlation-analysis

Soren, H. (2006). Example of Multivariate Analysis in R – Principal Component Analysis (PCA).

Usman, U., Yakubu, M. and Bello, A. Z. (2012). An Investigation on the Rate of Crime in Sokoto State Using Principal Component Analysis. *Nigerian Journal of Basic and Applied Science*. 20(2): 152-160

Wafula, M. E., Samson, W. W. and Chris, M. (2019). Application of Principal Component Analysis to Crime Data, Case Study: Mathare Slums, Nairobi County in Kenya. *American Journal of Theoretical and Applied Statistics*. 8(1), 7-17. https://doi.org/10.11648/j.ajtas.20190801.12