IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN:2319-765X. Volume 9, Issue 3 (Nov. – Dec. 2013), PP 08-17
www.iosrjournals.org

Detecting Factors impacting Crime Rates in Nigeria using Principal Component Analysis

H. G. Dikko[1], Abdulhameed A. Osi[2] and Yusif Bello[3]
[1]*Department of Mathematics, Ahmadu Bello University, Zaria Nigeria.*
[2]*State Criminal Investigation Department, Nigeria Police Katsina Command*
[3]*Department of Maths and Statistics, Hassan Usman Katsina Polytechnic, Katsina Nigeria*

Abstract: *Principal Component Analysis(PCA) is a data analysis tool that is used to reduce the dimensionality of a large number of interrelated variables while retaining as much of the information as possible. In this paper, PCA has been utilized on the crime data of Nigeria to discover the distinct influential variables; in addition which variables have silence in the identification of State being safe or dangerous. From the result, four Principal Components (PCs) have been retained using both scree plot and Kaiser's criterion which accounted for 75.024% of the total variation.*
Keywords: *Principal Component Analysis (PCA), crime data*.

I. Introduction

The scope of crime and the concern for its prevention/control has grown considerably in the last few years in Nigeria. What has previously been the predominant concern of the police and the private security industry had, in recent times spread to other groups, with interest in economic and even socio political aspects of our lives. The shopping malls, car dealers, real estate developers, financial institutes, news agencies, network service providers, international agencies, NGOs, residents groups and indeed, the general public are all concerned with the level/rate of crime in the Nigeria society. In the last four decades there has been an outstanding increase in criminal activities as some reports and studies have conferred in Nigeria [1].

There is no disagreement from both macro-and-micro level studies that the rate of crime in Nigeria has reached an unacceptable level. The fact file on losses between June 1999 and October 2001 painted a picture of robbery and murder victims akin to a declaration of war by hoodlums. Estimated property lost is in billions of Naira, while a total of 3,680 people lost their lives. Apart from those murdered in the course of robbery, murder occurs in a variety of circumstances. Of the 7,823 murders which occurred in Nigeria between 1989 and 1993, some started as arguments, fights, or brawls, while others were killed in murders, which began as robberies, rapes and burglaries. Very few were murdered in unknown circumstances[2]. A syndicate of young boys who specialized in killing commercial motorcyclists in the process of stealing their bikes confessed that they robbed either in the day or night by hiring commercial motorcycles and directing the cyclists to dark or lonely end of the town after enticing them with good pay. "As soon as he stops, one of us will put a cable around his neck and tighten the noose. While he is gasping for breath, we will drag him until he is dead. Then we will take his motorcycle away" [3].

However, what factors may be related to safety? Are there some variables you can observe to predict the safety and security of states even without using the actual crime rate statistics? We want to have a baseline to classify a particular State.

We try to solve the above questions with the help of multivariate statistical techniques based on the variables used to profile Nigeria States. We will use Principal Component Analysis to do pattern recognition, determining which variables greatly affect States' safety. The potential results are valuable for police department to increase enforcement and step up patrols in unsafe States; for State governments to decrease the crime rate by increasing or decreasing certain variables.

1.1 Crime

The word 'crime' is from a Latin word *crimen* meaning "accusation" or "fault". There are many definitions of crime. The definition employed by any particular scholar has to do more with his/her area of specialization. Different scholars give different explanations about the occurrence of crime. Who do we blame? Is it the individual who commits the act, the society or both the individual and the society? The question you need to ask is: Do we consider all acts of disobedience as crime? What are the components or elements you need to look when you mention the word 'crime'? When can you say crime has indeed occurred or has not? It is the view of one of the prominent scholars that any attempt to present a universal definition of crime may have some difficulties. This is because acts that are defined as criminal vary with time and place. An act may be criminal in

one society, but not in another. Likewise an act defined as crime at one time may not be at another. We can also say that the gravity or seriousness with which each society views the acts may be different.

Legal Definition of Crime

According to Tappan "crime is an intentional act in violation of the criminal law (Statutory and case law), committed without defence or excuse, and penalized by the state as a felony or misdemeanour" [4]. Crime is a violation of the criminal law, which is subsequently followed by legal punishment. A crime is an act or omission, which attracts sanctions. In studying the offender there can be no presumption that arrested, arraigned, indicted, or prosecuted persons are criminals unless they are also held guilty beyond all reasonable doubt of a particular offence. One thing you should bear in mind is that, one advantage of a legal definition of crime is that it is narrower and less ambiguous than a social definition of crime. Also, it is only a behaviour that violates the criminal law by definition that you can regard as a crime.

1.2 Classification of Crime

The classification of crime differs from one country to another. In the United States, the Federal Bureau of Investigation tabulates the annual crime data as Uniform Crime Reports (UCR). They classify violations of laws which derive from common law as Part I (index) crimes in UCR data, further categorized as violent or property crimes. Part I violent crimes include murder and criminal homicide (voluntary manslaughter), forcible rape, aggravated assault, and robbery; while Part I property crimes include burglary, arson, larceny/theft, and motor vehicle theft. All other crimes count as Part II crimes [5].

In Nigeria, the police classification of crimes also depends on what law prescribed. In Nigeria Police Abstract of Statistics (NPACS), offences are categorized in to four main categories:

- *Offences against persons* includes: manslaughter, murder and attempted murder, assault, rape, child stealing, grievous hurt and wounding, etc.
- *Offences against property* includes: arm robbery, house and store breakings, forgery, theft/stealing, etc.
- *Offence against lawful authority* include: forgery of currency notes, gambling, breach of peace, bribery and corruption, etc.
- *Offences against local act* include: traffic offences, liquor offences, etc.

1.3 Causes of Crimes

Criminal behaviour cannot be explained by a single factor, because human behaviour is a complex interaction between genetic, environmental, social, psychological and cultural factors. Different types of crimes are being committed by different types of people, at different times, in different places, and under different circumstances [2]. Here we discuss some of the causes of crime.

1. **Biogenetic factor**s: Criminologists are with the opinion that criminal activity is due to the effect of biologically caused or inherited factors. According to [6], a criminal is born, not made; that criminals were the products of a genetic constitution unlike that found in the non-criminal population.
2. **Social and Environmental factor** [7]: The environment is said to play significant role in determining criminal behaviour. Factors within the environment that mostly influence criminal behaviour includes poverty, employment, corruption, urbanisation, family, moral decadence, poor education, technology, child abuse, drug trafficking and abuse, architectural or environmental design [8] and [9] have attributed the current crime problem in Nigeria to urbanisation, industrialisation and lack of education. Kutigi [10] has said that the factors of crime in Nigeria are poverty and ignorance which are at the same time the opinion of many Nigerians [11]. In another dimension, according to [12], lack of integrity, transparency and accountability in the management of public funds, especially at all levels of government have been identified as the factors responsible for the endemic corruption that has eaten deep into the fabric of the Nigerian society over the years.

1.4 Statistics of Crimes in Nigeria

Nigeria has one of the highest crime rates in the world. Murder often accompanies even minor burglaries. Rich Nigerians live in high-security compounds. Police in some states are empowered to "shoot on sight" violent criminals [13].

In the 1980s, serious crime grew to nearly epidemic proportions, particularly in Lagos and other urbanized areas characterized by rapid growth and change, by stark economic inequality and deprivation, by social disorganization, and by inadequate government service and law enforcement capabilities [14].

Annual crime rates fluctuated at around 200 per 100,000 populations until the early l960s and then steadily increased to more than 300 per 100,000 by the mid-1970s. Available data from the 1980s indicated a continuing increase. Total reported crime rose from almost 211,000 in 1981 to between 330,000 and 355,000 during 1984-85. Although serious crime usually constituted the larger category, minor crimes and offences

accounted for most of the increase. Crimes against property generally accounted for more than half the offences, with thefts, burglary, and breaking and entering covering 80 to 90 percent in most years. Assaults constituted 70 to 75 percent of all offences against persons. The British High Commission in Lagos cited more than 3,000 cases of forgeries annually [14].

1.5 Multivariate Analysis in Criminology

Different researchers in criminology study different features relating to the concept, nature, prevention and control of crime. For instance, the location in space and time is an essential feature of any criminal event. The two components are so closely linked that they are considered simultaneously. This has led to so much focus upon spatial domain called the "hot spots" of crime [15].

Techniques such as Multivariate Analysis of Variance, Discriminant Function Analysis, Factor Analysis, Cluster Analysis [16], Log-Linear Analysis [17], Multi-Dimensional Scaling [18], and the more sophisticated Structural Equation Analysis [19] were used by criminal justice scholars in the examination of the crime phenomenon.

An Application of Multivariate Statistical Analysis on crime in U.S cities, analyzed 14 variables on 100 US cities using Principal Component Analysis, Factor Analysis and Discriminant Analysis to categorize the cities into two groups: high crime rate and low crime rate [20]

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. In [21] for 1977 US crime data, the overall crime rate was determined from the first PC, and the same result was achieved by [22] for the 1985 US crime data. The second PC which is interpreted as 'type of crime component' has successively classified the seven crimes into violence and property crimes. Also two researchs was conducted on th analysis of Crime data Using Principal Component Analysis: A case study of Katsina State [23] and An investigation on the Rate of Crime in Sokoto State using Principal Component Analysis [24].

In contrast to the deterrence theory which might lead to prediction that a longer period in detention might lead to a decrease rate of future offending, the review of the empirical literature by [25] on variance estimation using two population means, concluded that length of incarceration had no clear effect on the post release behavior. Also it was found by using Multiple Regression analysis that the faster the police respond to crime the lower the crime rate may be in a given community [26].

II. Materials And Methods

2.1 Data Collection and Exploration

When talking about factors which may have impact on the crime rate, everyone can think of several variables. Fortunately, variables affecting crime has been the investigation subjects of many discipline historically. Therefore, we can summarize those known factors to impacting crime volume and crime type occurring from place to place [27] as follows:-

1. Population density and level of urbanization
2. Composition of the population, particularly percentage of youth
3. Mode of transportation system
4. Economic condition
5. Cultural and Educational factors
6. Climate factors
7. Effective strength of law enforcement agencies
8. Citizens' attitudes toward crime

2.2 Description of Data

The data of poverty rate, unemployment rate, average household size, youth illiteracy rates and percentage moderate income household live hood was collected from National Bureau of Statistics (Annual Report 2013) , population, population density, sex ratio from the National Population Commission (2006 census), GDP(PPP) from ministry of finance, drug arrest and seizure arrest index(SAI) from National Drug Law Enforcement Agency (Annual report 2011), 2011 Presidential election percentage voters turnout for Independent National Electoral Commission (INEC) and police strength from Nigeria Police force. The dataset totally has 37 entries; each entry represents the information of a particular state in Nigeria. Table 3.1 gives entries of the dataset, Table 3.2 are the list of all the variables and what they stand for. It's worth mentioning that we do not include all the common variables listed above because some statistics of a States in Nigeria are not measurable, applicable or are difficult to find.

2.3 Principal Component Analysis

Principal component analysis (PCA) is probably the most popular multivariate statistical technique and it is used by almost all scientific disciplines. It is also likely to be the oldest multivariate technique. In fact, its origin can be traced back to Pearson (1901) or even Cauchy (1829, see Grattan-Guinness, 1997, p. 416), or Jordan (1874, and also Cayley, Silverster, and Hamilton, see Stewart, 1993; Boyer and Merzbach, 1989, for more) but its modern instantiation was formalized by Hotelling (1933) who also coined the term *principal component*. PCA analyzes a data table representing observations described by several dependent variables, which are, in general, inter-correlated. Its goal is to extract the important information from the data table and to express this information as a set of new orthogonal variables called *principal components*. PCA also represents the pattern of similarity of the observations and the variables by displaying them as points in maps (see, for more details Jolliffe, 2002; Jackson, 1991; Saporta and Niang, 2009).

Let the random vector $X' = \left[X_1, X_2, ..., X_p \right]$ have the covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq ..... \geq \lambda_p \geq 0$.

Consider the linear combinations

$Y_j = \ell_1^{'}\mathbf{X} = \ell_{j1}X + \ell_{j2}X + ... + \ell_{jp}X_p = \sum_{k=1}^{p} \ell_{jk}X_k$, j=1,2,…..p of thr element of X, where ℓ_j is a vector of p components $\ell_{j1}, \ell_{j2}, ..., \ell_{jp}$.

Then,

$$Var(Y_j) = \ell_j^{'}\Sigma\ell_j \qquad i = 1, 2, ..., p \qquad (1.1)$$

$$Cov(Y_j, Y_k) = \ell_j^{'}\Sigma\ell_k \qquad i, k = 1, 2, ..., p \qquad (1.2)$$

The PCs are those uncorrelated linear combinations $Y_1, Y_2, \ldots\ldots, Y_P$ whose variances in (1.1) are as large as possible. In finding the PCs we concentrate on the variances. The first step is to look for a linear combination $\ell_1^{'}X$ with maximum variance, so that

$$\ell_1^{'}X = \ell_{11}X_1 + \ell_{12}X_2 + ... + \ell_{1p}X_p = \sum_{k=1}^{p} \ell_{1k}X_k$$

Next, look for a linear combination $\ell_2^{'}X$ uncorrelated with $\ell_1^{'}X$ having maximum variance, and so on, so that at the k[th] stage a linear combination $\ell_k^{'}X$ is found that has maximum variance subject to being uncorrelated with $\ell_1^{'}X, \ell_2^{'}X, ..., \ell_{k-1}^{'}X$. The k[th] derived variable $\ell_k^{'}X$ is the k[th] PC. Up to p PCs 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 percentage of variance that any particular principal component accounts for can be calculated by dividing the variance of that component by the sum of all the variances, i.e.

$$\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i}$$

We use the high correlations between the principal components and the original variables to define which components we will utilize and which ones we will discard. One device that assists us in this decision process is a scree plot. Scree plots are graphs of the variance (eigenvalue) of each principal component in descending order. A point called an “elbow” is designated. Below this point is where the graph becomes somewhat horizontal. Any principal components whose variances lie above this point are kept and the others are discarded. The original variables that are highly correlated with each principal component that is kept determine what the label of that particular component will be.

III. Result And Discussion

2.1 Normality of the data

Often, before doing any statistical modeling, it is crucial to verify whether the data at hand satisfy the underlying distribution assumptions. Multivariate normal distribution is one of the most frequently made

distributional assumptions when using multivariate statistical techniques, e.g. Principal Component Analysis and Discriminant Analysis. Also, from an important property of multivariate normal distribution, we know that if $X=(X_1, X_2, ..., X_P)$ follow the multivariate normal distribution, then its individual components $X_1, X_2, ..., X_P$ are all normally distributed. Therefore, we need to examine normality of each Xi to guarantee that $X=(X_1, X_2, ..., X_P)$ is multivariate normal distributed.

Here, we use quantile-quantile plot (QQ plot) to assess normality of data though there are more formal mathematical assessment methods. The reason is that with a large dataset, formal test can detect even mild deviations from normality which actually we may accept due to the large sample size. However, a graphical method is easier to interpret and also have the benefit to easily identify the outliers. In QQ plot, we compare the real standardized values of the variables against the standard normal distribution. The correlation between the sample data and normal quantiles will measure how well the data is modeled by a normal distribution. For normal data, the points plotted should fall approximately on a straight line in the QQ plot. If not, data transformation like logarithm, square root and power transformation can be applied to make the data appear to more closely normally distributed.

Drawing QQ plot of each variable reveals that 6 out of the 13 variables approximately follow normal distribution, they are: MODERATE, POVERTY, UNEMPLOYMENT, VOTERS, SEXRATIO and AVRG. We try different forms of transformation on the remaining 7 variables to obtain the substitute variables which perform better on normality. Table 3.1 list the specific forms used to get the new variables. Figure 3.1 is the QQ plot of POP and GDP in terms of both before-transformation and after-transformation. We can see the effectiveness of data transformation method because data tends to be normally distribution after transformation.

New Variables
LOG(POP)
LOG(DENS)
LN(ILLITRACY)
MODERATE
POVERTY
UNEMP
SEXRATIO
AVERAGE
VOTERS
SQRT_LOG(POLICE)
LN_LN_LOG(GDP)
LOG(ARREST)
LOG(SAI)

Table: 3.1

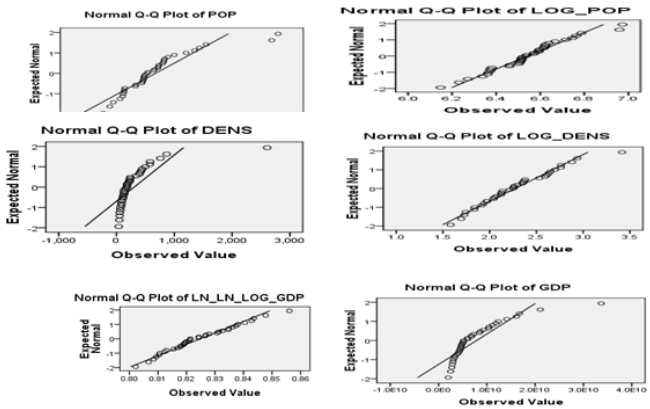

Figure 3.1: Normality check samples

Figure 3.1: Normality check samples

	POP	**DEN**	**SEX**	**AVRG**	**ILLT**	**MOD**	**PVR**	**UNR**	**VOT**	**POL**	**GDP**	**ARRT**	**SAI**
POP	1.00	0.407	-0.003	0.046	-0.005	0.107	-0.017	-0.016	-0.030	-0.459	0.665	0.524	0.096
DENS	0.407	1.000	-0.014	-0.483	-0.755	-0.080	-0.366	-0.414	0.158	-0.155	0.603	0.462	0.039
SEX	-0.003	-0.014	1.000	0.258	0.000	0.008	-0.339	-0.065	0.276	0.276	0.118	0.202	-0.124
AVRG	0.046	-0.483	0.258	1.000	0.643	-0.163	0.533	0.522	0.249	-0.149	-0.388	0.042	-0.437
ILLT	-0.005	-0.755	0.000	0.643	1.000	0.265	0.574	0.473	-0.205	-0.118	-0.500	-0.163	-0.131

MOD	0.107	-0.08	0.008	-0.163	0.265	1.000	-0.100	-0.089	-0.623	0.209	0.007	0.037	0.300
PVR	-0.017	-0.366	0-.339	0.533	0.574	-0.100	1.000	0.370	0.001	-0.465	-0.457	-0.125	-0.198
UNR	-0.816	-0.414	0-.065	0.522	0.473	-0.089	0.370	1.000	0.291	-0.097	-0.244	-0.156	-0.297
VOT	-0.03	0.158	0.276	0.249	-0.205	-0.623	0.001	0.291	1.000	-0.132	0.097	0.015	-0.521
POL	-0.459	-0.155	0.276	-0.149	-0.118	0.209	-0.465	-0.097	-0.132	1.000	0.106	-0.096	0.236
GDP	0.665	0.603	0.118	-0.388	-0.500	0.007	-0.457	-0.244	0.097	0.106	1.000	0.441	0.353
ARRT	0.524	0.462	0.202	0.042	-0.163	0.037	-0.125	-0.156	0.015	-0.096	0.441	1.000	-0.138
SAI	0.096	0.039	-0.124	-0.437	-0.131	0.300	-0.198	-0.297	-0.521	0.236	0.353	-0.138	1.000

Determinant = 9.84E-005

Table 3.2: Correlation Matrix for the variables

The correlation matrix on table 3.2 has displayed different levels of correlation between the variables. The correlation coefficient is between -1 and 1. A coefficient of 1 means perfect positive correlation, -1 means a perfect negative correlation and 0 means no correlation. Looking at the table there are very low correlations between population on one hand and sex ratio, illiteracy rate and average household size also between population density and drug seizure arrest index and moderate household income but a high correlation with Illiteracy rate, there is no correlation between sex ratio and illiteracy rate but very low correlation with moderate household income. A very low correlation between average household size and number of drug offenders arrested but high correlation with illiteracy rate. And the determinant of correlation matrix equal to 9.84E-005 (0.000984) which is greater than the necessary value of 0.00001. Therefore, multicollinearity is not a problem for this data there is no need to consider eliminating any variable at this stage.

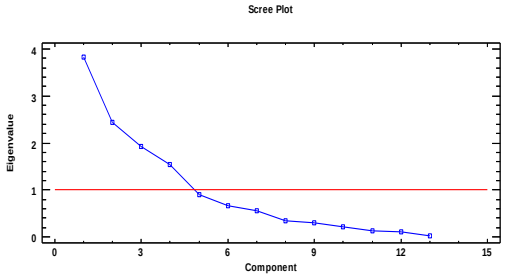

Figure 3.2: The Scree Plot of the PCs

Component Number	*Eigenvalue*	*Percent of Variance*	*Cumulative Percentage*
1	3.83184	29.476	29.476
2	2.44657	18.820	48.295
3	1.92905	14.839	63.134
4	1.54566	11.890	75.024
5	0.890268	6.848	81.872
6	0.664219	5.109	86.982
7	0.547749	4.213	91.195
8	0.352838	2.714	93.909
9	0.291824	2.245	96.154
10	0.224953	1.730	97.884
11	0.139064	1.070	98.954
12	0.103291	0.795	99.749
13	0.0326696	0.251	100.000

Table 3.3: Eigenvalues of the Correlation Matrix

The first and very important step of PCA is to determine the number of PCs. Here, we use the Kaiser's criterion to do the selection which is accurate when there are less than 30 variables and communalities after extraction are greater than 0.7 . Therefore based on Kaiser's rule we retain all PCs with eigenvalues greater than 1, which leave us with first four PCs that explain up to 75.024 percent of the total variability of the data set as shown in the Table 3.3.

We can also use the scree plot to choose the number of PCs to retain, as shown in Figure 3.2 the "elbow" appears at the 5th PC, therefore we should keep the first four PCs which account for 75.024 % of the total variance , as a valuable reduction in dimensionality.

Therefore, it is a fact that the first four PCs accounted for 75.024% of total variance of the original variables and simultaneously reduce the data dimension from 13 to 4.

	Component	*Component*	*Component*	*Component*
	1	2	3	4
POP	0.183035	-0.298347	0.513504	0.145955
DENS	0.405331	-0.245448	0.041937	-0.153541
SEXRATIO	0.0399517	-0.124233	-0.260236	0.616974
AVERAGE	-0.378038	-0.23778	0.0662794	0.30599
ILLITRACY	-0.408727	0.0974004	0.275208	0.245502
MODERATE	0.0452321	0.374277	0.327189	0.303141
POVERTY	-0.359101	-0.0993719	0.289253	-0.238104
UNEMP	-0.323175	-0.15862	0.0310507	0.0994551
VOTERS	-0.0500998	-0.474311	-0.373782	-0.0262072
POLICE	0.0901235	0.324002	-0.365567	0.375952
GDP	0.399922	-0.165723	0.148844	0.163808
ARREST	0.203726	-0.300049	0.2519	0.304982
SAI	0.207662	0.383047	0.189885	-0.01455

Table 3.4: Table of Component Weights

The next step is to look at the content of variables that load onto the same factor to try identifying common themes. In Table 3.4, we visualize the coefficients of the variables in each PC and combined with the exact values of these coefficients (also known as loading matrix of PCA), we can clearly see which variables are dominant in each PC. For example in PC2, SAI, MODERATE and ARREST have relative large absolute value, hence we denote PC2 as drug abuse factor. Similarly, view PC3 as population factor, PC4 as factor related to police strength. However, it is kind of difficult to define PC1 in which ILLITRACY, AVERAGE, DENSITY, UNEMPLOYMENT, POVERTY and GDP all have relative large absolute value of greater than 0.3 level, which indicates interpretation of PCs is one major disadvantage of PCA.

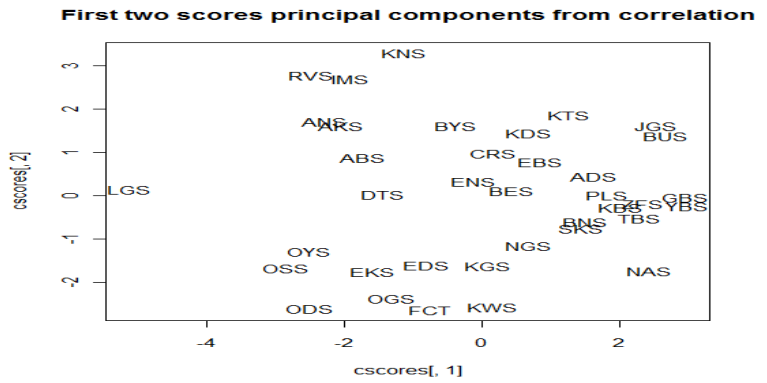

Figure 3.3a score plot of first and second PCs

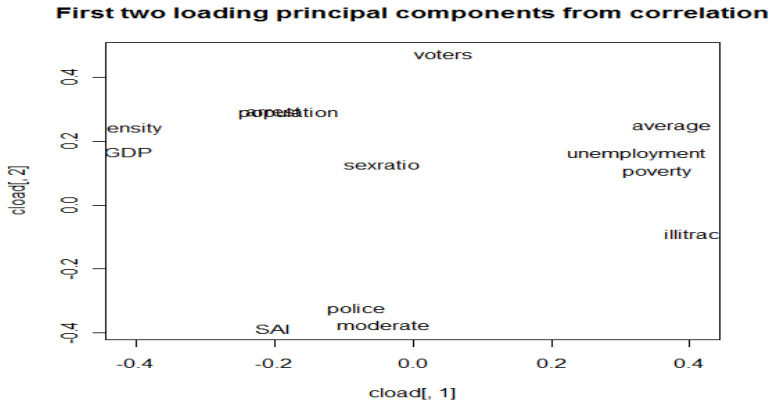

Figure 3.3b Loading plot of first and second PCs

From Fig. 3.3a and Fig. 3.3b we can see from loading and scores plot the State with combine prevalence of Population Density and GDP is Lagos State (LGS).

Bauchi (BAU), Kaduna (KDS) and Jigawa (JGS) States have higher Average household size, while Zamfara (ZFS), Bauchi (BUS), Niger (NGS) and Gombe (GBS) States with highest Unemployment Rates.

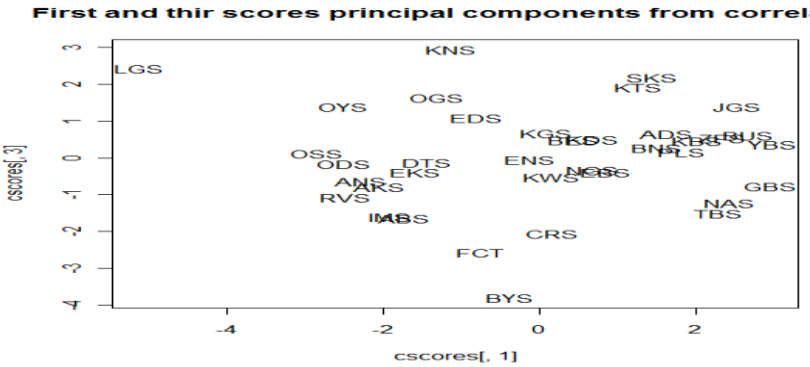

Figure 3.4a Scores plot of first and third PCs

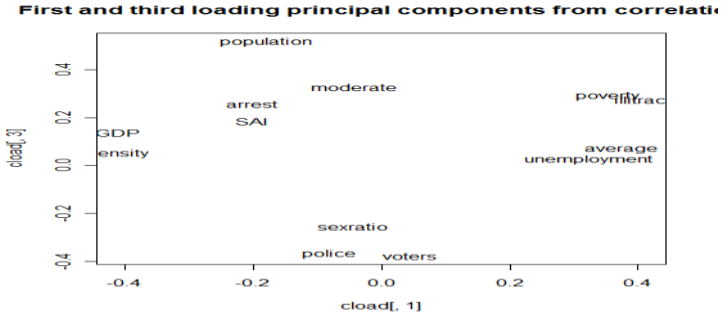

Figure 3.4b Loading plot of first and third PCs

Fig. 3.4a and Fig. 3.4b and figure 4.5b shows that Kano (KNS) and Lagos (LGS) States are the most populous States, while Sokoto (SKS), Katsina (KTS), Gombe (GBS), Jigawa (JGS), Bauchi (BUS) and Adamawa (ADS) States have the predominance of Poverty Rates

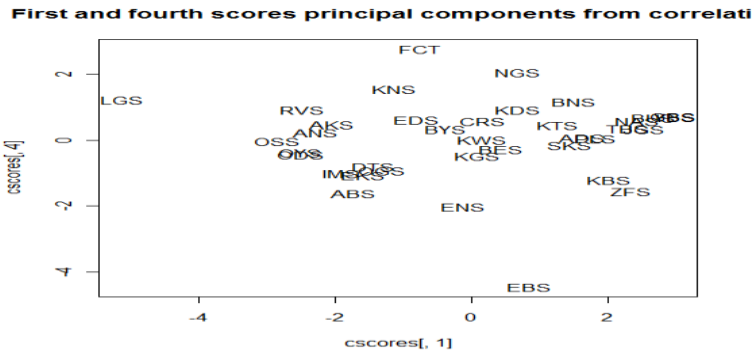

Figure 3.5a Scores plot of first and fourth PCs

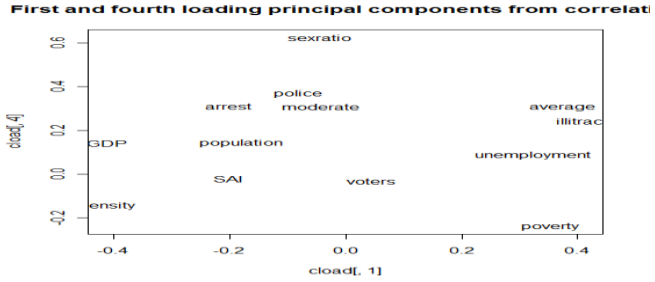

Figure 3.5b Loading plot of first and fourth PCs

Fig. 3.5a and Fig. 3.5b show that FCT Abuja has the higher male to female Sex ratio

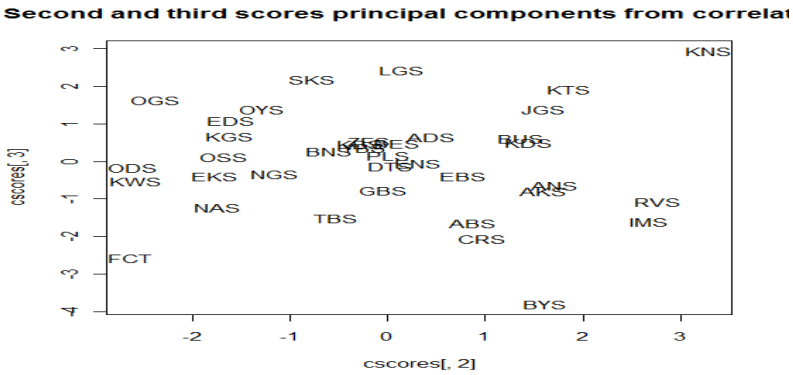

Figure 3.6a Scores plot of second and third PCs

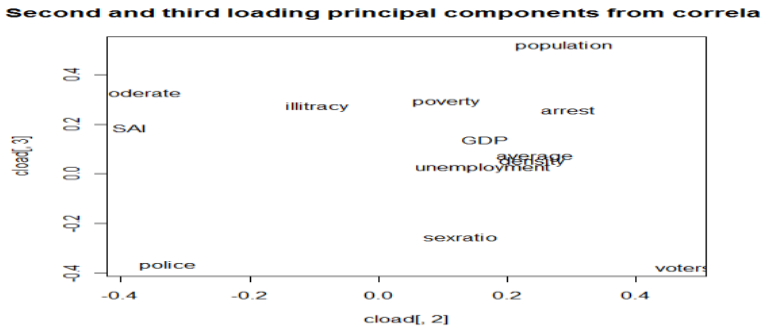

Figure 3.6b Loading plot of second and third PCs

Figure 3.6a and 3.6b shows Kano (KNS), Katsina (KTS) and Jigawa (JGS) States have a tendency towards number of drug offender arrested, while Bayelsa (BYS), Abia(ABS), Cross River (CRS), Imo (IMS) and Rivers States (RVS) with higher turnout for 2011 presidential election, at the same time Ondo (ODS) State has greater quantity of drug Seizure Arrest Index.

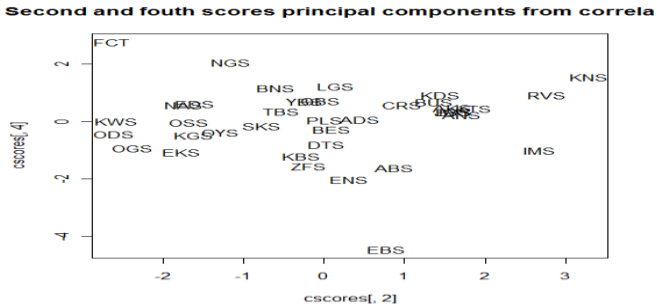

Figure 3.7a Scores Plot of second and fourth PCs

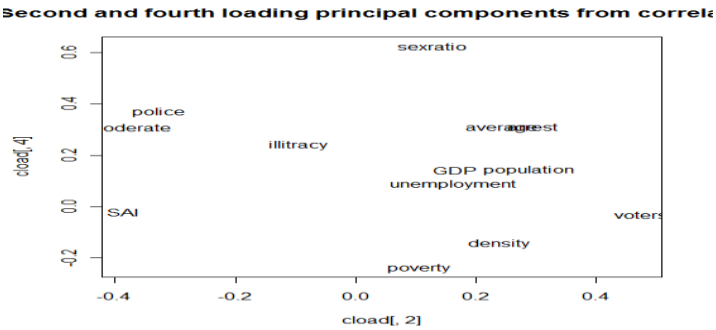

Figure 3.7b Loading Plot of second and fourth PCs

From Fig. 3.7a and Fig. 3.7b FCT Abuja and Nasarawa (NSR) State have high proportion of number of police officers per 100,000 populations.

IV. Conclusion

Principal Component Analysis is successfully applied into our data by extracting four PCs out of the 13 original variables, which implies a great dimensionality reduction. In addition, these PCs account for almost 75.024% variance of the original dataset, thus we do not loss much information.

Highly correlated original variables with a particular PC can serve to determine the label of that PC. Illiteracy rates, average household size, population density, unemployment rates and poverty rates collaboratively make contribution to PC1. PC2, PC3, and PC4 represent drug trafficking and abuse factor, population factor and a factor related to gender and police strength respectively. That is to say, the above nine variables can considerably help when classifying a community as safe or unsafe.

Our analysis has established that Illiteracy rates, population density, unemployment rates and poverty rates have impact on crime, States with high illiteracy rate, unemployment rates, poverty rates and densely populated state is expected to have high crime rates. Secondly population, police strength and drug trafficking and abuse also have impact on crime; as the lower the number of police personnel in the state the low the effect of crime control, also the higher the number of people in the state the crime rates.

It is therefore recommended that government should double its efforts to reduce the levels of poverty, unemployment and illiteracy through better funding and actualization small scale industries, invest in Agriculture and Education as well. Secondly government should employ more police personnel as this would first, provide more hands in crime management and secondly reduce unemployment, finally Government should also address the drug abuse issues especially for youth

With help of the Multivariate Method of PCA we reduce the 13^{th} variables that can affect the crime rate of the state to four important variables that show a high correlation with all the thirteen variables. PCA did decrease the number of variables to four and accounted for 75.024% of the total variation. The three PCs give us a good summarization of sample variation and reduction in the data from 37 by 13 to 37 by 4 PCs; therefore we can perform further analysis, namely Discriminant Analysis on these PCs and original variables to classify the States as safe and unsafe.

References

[1] Cleen (2007). The Nigerian Police as at November 2007. http://www.cleen.org/sumarry%20crime% 20statistics% 20in%20nigeria%202007.p

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

[3] Ifeanyi (2004) Nigeria Crime Statistics

[4] Tappan (1964:32) Definition of crime

[5] Wiki/cr (2009). Crime, http://en.wikipedia.org/wiki/criminal_94.

[6] Lombroso, C. (1911). The Criminal Man, New York: Putman.

[7] Sutherland, E.H. (1939). The *white-colar criminal. American Sociological Review*, **5:** *1-12*

[8] Oyebanji, J.O. (1982). *Economic Development and the Geography of crime: an Empirical Analysis.*

[9] Akpan, A.U. (2002). *Notion of causes of crime among Nigerians. International Journal of Social and Policy Issues:* **1(1):** *36-40*

[10] Kutigi, I. (2008). Puberty, Ignorance and Evil Intentions as strong crime factors. Tribune, Monday 23/06/2008.

[11] Azubuike, C. (2007). Crime and control strategies Daily Sun, Tuesday 10/2/2009.

[12] Ayoola, A. O. (2008). Ayoola identifies causes of corruption. This Day, 20/2/2008.

[13] Financial Times (2009). London, Thursday 30/04/2009.

[14] Nigeria (1991). Nigeria Incidence of crime rate. http://www.photius.com/countries/

[15] Sherman, L. W., P. R. Gartin and M. E. Buerger. 1989. "Hot Spots of Predatory Ccrime: Routine Activities and the Criminology of Place." *Criminology **27**: 27-55*

[16] Mc Gurk *et al* (1981) Log-linear analysis

[17] Crowly Joan E. and Lynne M. Adrian. 1992. "Women Misdemeanants in the Allegheny County Jail 1892-1923". *Journal of Criminal Justice 20:311-331*

[18] Smith, William R.; Smith, Randall D. and Noma, Elliot. 1986. "The Multidimensionslity of Crime: A Comparison of Techniques for Scaling Delinquent Careers" *Journal of Quantitative Criminology 2 (4):329-352*

[19] Liu, Xiaoru and Howard B. Kaplan. 1999. "Explaining the Gender Difference in Adolescent Delinquent Behavior: A Longitudinal Test of Mediating Mechanisms." *Criminology 37 (1):195-215*

[20] Williams, K and Gedeon, R (2004); Howard University http://www.users.muohio.edu/porterbm/sumj/2004/CrimeStats.

[21] Printcom (2003), http://support.sas.com/onlinedoc/912/getDoc/common.hlp/images/ copyrite.htm.

[22] Hardle, W. and Zdenek, H. (2007). Multivariate Statistics: Exercises and Solutions. *New York: Springer-verlag.*

[23] Shehu U. Gulumbe, H. G. Dikko and Yusuf Bello "Analysis of Crime Data using Principal Component Analysis: A case study of Katsina State" *CBN Journal of Applied Statistics*

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

[25] Hood Roger and Richard Sparks (1990) Key issues in Criminology. *New York Mc Graw Hill*

[26] Kohfeld Carol W., and John Sprague (1990): Demography, Police behavior and Deterrence. *Criminology 28:11-136*

[27] U.S Department of Justice, "Variables Affecting Crime", *Crime in the United States, 2009, 2010.*

[28] Hardle, W. and Simar, L. (2003). Applied Multivariate Statistical analysis, 2nd edition. *New York: Springer-verlag.*

[29 Jolliffe, I.T. (2002). Principal Component Analysis 2nd edn. *New York: Springer-verlag*.

[30] Lindsey Smith (2002): A tutorial on Principal Component Analysis, http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components .pdf

[31] Rencher, A.C (2002). Methods Multivariate Analysis, 2nd edn. *New York: Wiley*.

[32] Richard, A.J and Dean, W.W (2001). Applied Multivariate Statistical Analysis, 3rd edn. *New Delhi: Prentice-Hall.*