Analyzing the Distribution of Crimes in Oyo State (Nigeria) using Principal Component Analysis (PCA)

Olufolabo.O.O*, Akintande O.J*, Ekum M.I**

Department of Statistics, Yaba College of Technology Department of Mathematics, University of Lagos Department of Mathematics, University of Lagos

Abstract: This paper analyses Oyo State crime data which consists of eighteen major crimes reported to the police between the periods of 1996 – 2014. The crimes consist of robbery, kidnapping, house and store breakings, theft/stealing, grievous hurt and wounding, murder, rape, assault etc.

There is an average of 56 cases of Murder, 152 cases of Armed Robbery, 12 cases of Suicide, 541 cases of Grivious Harms/Wounding, 1196 cases of Assaults, 392 cases of Burglary/Stealing, 263 cases of Store Breaking, 544 cases of House Breaking, 1385 cases of Stealing/Theft, 99 cases of Rape/Indecent Assaults, 12 cases of Kidnapping, 442 cases of False Pretence/Cheating, 110 cases of Unlawful Possession, 16 cases of Arson, 327 cases of Breach of Peace, 23 cases of Forgery, 8 cases of Child Stealing and 29 cases of Unnatural Offences.

Correlation analysis and principal component analysis (PCA) were employed to explain the correlation between the crimes and to determine the distribution of the crimes in the state. The result shown a significant correlation between Store breaking and Stealing. Assaults have a strong positive relationship with False Pretence and Breach of Peace. Kidnapping has a strong positive relationship with False Pretence Breach of Peace. False Pretence has a strong positive relationship with Breach of Peace. There is weak positive relationship between Murder case and Armed Robbery, Suicide, Grivious harm/wounding, Assaults, Store breaking, House breaking, Stealing, False Pretence, Unlawful possession, Breach of peace, Forgery, Child stealing and Unusual crimes. Murder has a positive relationship with Rape and Arson. And a weak negative relationship with Burglary. Armed Robbery has a positive relationship with Kidnapping and weak positive relationship with other forms of crimes, a weak negative relationship with Burglary, Store breaking, Stealing and Unlawful possession. Suicide has a positive relationship with Rape. Grivious harm/Wounding has a positive relationship with Assaults. Stealing/Theft is more prevalent in Oyo State. Besides Stealing/Theft, Assaults is prominent among other crimes. The PCA has suggested retaining six components that explain about 83.79 percent of the total variability of the

data set.

I. Introduction

Crime is a known word peculiar to every nations, societies and communities around the world. An adage in Yoruba says; "Ilu ti kosofin; ese osi nibe" This means that, there is no crime; where there is no law! We can infer therefore that, where there is law, there is crime! Law and order is a defined ways by which nation's controls crimes in the society. Crime is a menace which requires close monitoring and analysis. Hence, there is need to study the rate at which crimes are committed in the society and ways to curbing it. It is impossible to eradicate crime in the society but it is important to provide measures in curbing it. Despite the fact that, crime is inevitable in a society Durkheim (1933), various controlling and preventive measures had been taken, and are still being taken to reduce the menace.

Crime does not have any boundary. It has nothing to do with a country being developed, developing or under-develop. In fact, worst crimes are reported daily in the news in most developed countries of the world. Crime therefore varies across continents and it prevalence can be nation's characterized. We mean that, some crimes are peculiar with some nations, communities and societies. For instance, Kidnapping, Homicides, armed robbery and killing involving the use of armed weaponry are peculiar in most advanced countries. Suicide bombing, Rape, Homicides, Religion crime and Massacre are peculiar in the Arab nations. Kidnapping, Ethnic fight, Tribal war, Corruption, Rioting, Religion clashes etc., are peculiar in Africa, etc.

Our study focus only on a State (Oyo State) in Nigeria. In categorizing crimes, crime do varies even within States of a nation. For instance, in Nigeria today (2003-till date), Bombing, Terrorism, Suicide bombing, Religion war& killings, vandalism etc., is a crime peculiar in the Northern part of Nigeria particularly in the states like Borno, Adamawa, Kano and Yobe whereas, Kidnapping, Pipeline vandalism, Armed robbery and Oil theft are order of the day in Niger-Delta part of Nigeria. In the South west and in the Eastern part of Nigeria, we have heard of cases of Ritual killing, Theft &Stealing, Kidnapping, Assaults, Wounding, Armed robbery etc.

This implies that, Crime prevalence may varies across borders but there is no society, community or nation that is void of Crime.

Therefore, in Ibadan, Oyo state, there are 18 major categories of crimes associated with the people. Police department in the identified Stealing/Theft as the most committed crime in the state.

Several crime aid factors in societies includes but not limited to illiteracy, poverty, broken home, bad company, environment, religion belief and extremism, failure of Police and authority in administering justice. Just as crimes varies across the globe; it motivating factors also varies depending on the nation. In most advanced nations where the standard of living is considerably well enough, one may overrule Illiteracy and poverty as a crime motivating factor. What motivate crime in country **A** may be less significant in country **B**. In analyzing our own situation for this study, all these factors are significant relevant to crime rates in most of our societies.

II. Methodology

In order to test for similarities between crimes and the differences in the crimes committed in Oyo state over the last 20 years and to determine which of these crimes is/are significant to another.

We performed some descriptive review of the data in order to explain the distributions in the crime rates over the last 20 years in Oyo state. In other words, to determine the crime rates on yearly bases and to understand the prevalence of a crime over the other. The Principal Component Analysis helped us to go deeper into analyzing the significant-components of crime rates in Oyo state.

III. Principal Component Analysis.

Principal components analysis (PCA, for short) is a variable reduction technique that shares many similarities to exploratory factor analysis. Its aim is to reduce a larger set of variables into a smaller set of 'artificial' variables, called 'principal components', which account for most of the variance in the original variables. If these variables are highly correlated, it implies that the components extracted have a high degree of relationship with the original variables. This helps you understand whether some of the variables you have chosen are not sufficiently representative of the construct you are interested in, and should be removed from your new measurement. PCA is conceptually different to factor analysis, in practice it is often used interchangeably with factor analysis, and is included within the 'Factor procedure' in SPSS.

Methods:

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 $X' = [X_1, X_2, ..., X_p]$ have the covariance matrix \sum with eigenvalues $\lambda_1 \ge \lambda_2 \ge \lambda_3, \ldots, \ldots, \ge \lambda_p \ge 0$.

Consider the linear combinations,

 $Y_j = \alpha'_j X = \alpha'_{j1} X_1 + \alpha'_{j2} X_2 + \dots + \alpha'_{jp} X_p = \sum \alpha'_{jk} X_k, j = 1, 2, 3, \dots, p \text{ of the element of } X$, where α_j is a vector of p components $\alpha'_{j1}, \alpha'_{j2}, \dots, \alpha'_{jk}$

Then

 Var $(Y_j) = \alpha'_j \sum \alpha_j$ j = 1, 2, ..., p (1.1)

 Cov $(Y_j) = \alpha'_j \sum \alpha_k$ j = 1, 2, ..., p (1.2)

The PCs are those uncorrelated linear combinations Y1,Y2,...,Yp whose variances in (1.1) are as large as possible (Richard and Dean, 2001). In finding the PCs we concentrate on the variances. The first step is to look for a linear combination α'_1X with maximum variance, so that

 $\alpha'_{1}X = \alpha'_{11}X_{1} + \alpha'_{12}X_{2} + \dots + \alpha'_{1p}X_{p} = \sum \alpha'_{1k}X_{k}$

Next, look for a linear combination $\alpha'_2 X$ uncorrelated with $\alpha'_1 X$ having maximum variance, and so on, so that at the kth stage a linear combination $\alpha'_k X$ is found that has maximum variance subject to being uncorrelated with $\alpha'_1 X$, $\alpha'_2 X$,..., $\alpha'_{k-1} X$. The kth derived variable $\alpha'_k X$ is the kth PC. Up to p PCs could be found, but we have to stop after the qth stage (q ≤ 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 $(Y_j) = \alpha'_j \sum \alpha_j$ j = 1, 2, ..., p
- The total variance in a data set is equal to the total variance of PCs

 $\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum Var(X_j) = \lambda_1 + \lambda_2 +, \dots, + \lambda_p = \sum Var(Y_j)$

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

$$Z = \left[z = \frac{(X_j - \mu j)}{\sqrt{\sigma_{11}}} \right] \qquad j = 1, 2, \dots, p$$

In matrix notation

$$Z = V^{1/2} (X - \mu)$$

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 our previous properties for X are applied for the Z, so that the notation Yj refers to the jth PC and $(\lambda_{j}, \alpha_{j})$ refers to the eigenvalue – eigenvector pair. However, the quantities derived from Σ are not the same from those derived from ρ (Richard and Dean, 2001).

The jth PC of the standardized variables $Z=[z_1, z_2, ..., z_p]$ with Cov (Z), is given by

$$X_j = \alpha'_j Z = \alpha'_j (V^{1/2})^{-1} (X - \mu)$$

So that,

 $\sum Var(Y_j) = \sum Var(Z_j) = P \qquad j = 1, 2, \dots, p$ In this case, (λ_1, α_1) , $(\lambda_2, \alpha_2), \dots, (\lambda_p, \alpha_p)$ are the eigenvalue- eigenvector pairs for ρ with $\lambda_1 \ge \lambda_2 \ge \lambda_3, \dots, \lambda_p \ge 0.$

Interpretation of the Principal Components:

The loading or the eigenvector $\alpha_j = \alpha_1, \alpha_2, \alpha_3, \dots, \alpha_j$, is the measure of the importance of a measured variable for a given PC. When all elements of are positive, the first component is a weighted average of the variables and is sometimes referred to as measure of overall crime rate. Likewise, the positive and negative coefficients in subsequent components may be regarded as type of crime components (Rencher, 2002 and Printcom, 2003). The plot of the first two or 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 observations Yij are obtained simply by substituting the original variables Xij into the set of the first q PCs. This gives $Y_{ij} = \alpha'_{j1}X_{i1} + \alpha'_{j2}X_{i2} + \alpha'_{j3}X_{i3} + \dots + \alpha'_{jp}X_{ip}$ j = 1, 2, 3..., p; i = 1, 2.., n

The plot of the first two or three PCs against each other enhances visual interpretation (Soren, 2006).

The proportion of Variance:

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 graphical representation of the ability of the PCs to explain the variation in the data (Cattell, 1966).



IV. Analysis And Result Table 1: Crime rates across the year (1996-2014). Fig: 2.0: Crime Chart (1996-2014).

Source: Federal Department of State Police, Oyo State

We take 1996 as the base year for this analysis. In 1996; 4810 crimes were recorded. In 1997; 7061 were recorded. This amount to 2251 increase in the number of offenders. It amount to 46.8% increase in crime rate within a year.

Increase/Decrease rate $(I/DR) = \frac{|\text{Number of Current Crime -Number of Previous Crime |}}{\text{Number of Crime in the Previous Year}} X 100$ Between 1997 and 1998; the rate decreased to 5256 crimes. This amounts to 1805 decrease in the number of offenders. 25.6% decrease in crime rate. In 1999, the crime rates increased to 5548. This amount to 292 increase in crime rate.

In 2000; a total of 4476 crimes were recorded. This amount to 1072 decrease in crimes from the previous year (1999). It amount to 19.3% decrease in crime rates. In 2001; a total of 4789 crimes were recorded. This amount to 313 increase in crimes. This gives 6.99% increase in crime rates.

In 2002; a total of 5920 crimes were recorded. This amount to 1131 additional increase in the number of crimes. It resulted to 23.6% increase in crime rate. In 2003; a total of 4620 crimes were recorded. It amount to 1291 decrease in the number of offences. It amounts to 21.8% decrease in crime rate.

Year 2004 recorded the highest number of crime committers, 10803 offenders were booked. This amount to a sparring increase of 6174 additional offenders from the previous year (2003) and amount to 133.4% increase in crime rate. We cannot say precisely what caused this unusual increase in crime rate in Oyo state in 2004 but we sensed that some factors could be responsible.

The lowest crime rates is recorded in 2010. Only 1296 offenders were booked. This amount to 77.05% decrease in crime rates compare to the previous year (2009). In 2011; an unusual increase of 217.3% increase in crime rates was recorded compared to 2010.

2012, 2013 and 2014 have a total 4334, 6617 and 4540 crimes recorded respectively. And it amounted to 5.4% and 52.7% increase in crime rates in 2012 and 2013 respectively and 31.4% decrease in 2014.

	м	A	s	G	A	В	s	н	ST	R	ĸ	P	UN	AR	BR	F	с	U
м	1.00																	
A	0.47	1.00																
5	0.36	0.29	1.00															
G	0.40	0.28	0.24	1.00														
A	0.47	0.24	0.06	0.60	1.00													
В	-0.07	-0.41	-0.01	0.11	0.19	1.00												
S	0.46	-0.03	0.44	0.44	0.43	0.23	1.00											
H	0.42	0.12	-0.15	0.16	0.23	-0.04	0.09	1.00										
57	0.48	-0.08	0.33	0.46	0.39	0.34	0.80	0.14	1.00									
R	0.61	0.25	0.58	0.55	0.47	0.26	0.68	0.50	0.71	1.00								
K	0.52	0.50	0.09	0.40	0.55	-0.29	0.11	0.38	0.27	0.44	1.00							
Ρ	0.35	0.32	0.06	0.41	0.70	0.03	0.20	0.40	0.23	0.51	0.78	1.00						
UN	0.27	-0.11	-0.04	0.27	0.19	0.38	0.55	0.35	0.38	0.46	-0.20	0.11	1.00					
AR	0.56	0.11	0.13	0.06	0.32	-0.22	0.37	0.41	0.40	0.35	0.47	0.34	-0.05	1.00				
BR	0.37	0.37	-0.10	0.42	0.79	-0.06	-0.02	0.23	0.13	0.22	0.76	0.73	-0.16	0.23	1.00			
F	0.38	0.33	-0.24	0.24	0.37	0.33	-0.24	0.19	0.07	0.15	0.47	0.35	-0.06	0.08	0.60	1.00		
с	0.05	0.15	-0.24	-0.05	0.17	0.17	0.00	-0.07	0.05	-0.02	0.27	0.12	-0.30	0.21	0.21	0.51	1.00	
U	0.50	0.30	0.10	0.42	0.55	0.23	0.02	0.24	0.37	0.41	0.64	0.63	-0.03	0.35	0.71	0.71	0.07	1.00

Table 2.0: Correlation Matrix

In summary, Store breaking and Stealing has a very strong positive relationship. Assaults have a strong positive relationship with False Pretence and Breach of Peace.

Kidnapping has a strong positive relationship with False Pretence Breach of Peace. False Pretence has a strong positive relationship with Breach of Peace.

The strong or positive relationship implies that as one increases the other also increases. Murder has a positive relationship with Rape for instance, implies that, as murder cases increases, Rape cases increases. This means that, some Murder cases could be as a result of Rape, Arson, Armed Robbery, Suicide, Grivious harms/Wounding, Assaults, Store breaking, House breaking, Stealing, False pretence, Unlawful possession, Breach of peace, Forgery, Child stealing and Unusual crimes. That is all these crimes are correlated, hence, one may leads to another.

Principal Component Analysis (PCA). Toble 3.0: KMO and Bartlett's Test

Table 3.0: KMO and Bartlett's Test						
Kaiser-Meyer-Olkin Measure	.206					
	Approx. Chi-Square	327.500				
Bartlett's Test of Sphericity	df	153				
	Sig.	.000				

Although the KMO value is small (< 0.5) but the Bartlett's test of Sphericity indicates we can proceed. That is the test is highly significant.



Fig 1: Screen Plot of the Principal Components.

The eigenvalues and the cumulative proportions of the explained variance are displayed in Table 4. Considering the eigenvalue-one criterion and the Scree plot in figure 1, it would be reasonable to retain the first six PCs. A commonly accepted rule says that it suffices to keep only PCs with eigenvalues larger than 1. However, the seventh and the eighth eigenvalues and are approximately close to 1, so that the first 6 PCs can be retain to explain 83.79 percent of the total variability or the first 8 PCs to explain up to 91.90 percent of the total variability.

CRIMES	Eigenvalues	Proportion	Cumulative
MURDER	6.284	34.910	34.910
ARMED ROBBERY	3.139	17.439	52.349
SUICIDE	1.944	10.798	63.148
GRIVIOUS HARM/WOUNDING	1.384	7.691	70.838
ASSAULTS	1.304	7.242	78.081
BURGLARY/STEALING	1.027	5.708	83.788
STORE BREAKING	.834	4.634	88.423
HOUSE BREAKING	.626	3.479	91.902
STEALING/THEFT	.481	2.670	94.571
RAPE/INCEDENT ASSAULT	.367	2.039	96.610
KIDNAPPING	.252	1.397	98.007
FALSE PRETENCE/CHEATING	.172	.956	98.964
UNLAWFUL POSSESSION	.085	.471	99.435
ARSON	.044	.244	99.679
BREACH OF PEACE	.032	.177	99.856
FORGERY	.024	.136	99.992
CHILD STEALING	.001	.007	99.999
UNNATURAL OFFENCE	.000	.001	100.000

 Table 4.0: Eigenvalues

This table 5 below; concentrated on the six PCs that explains 83.788 per cent of the total variability of the data set are retained.

Component 1 has a positive relationship (weak and strong) with all the crimes recorded but majorly it identify (strong relationship with) Murder, Grivious harm/wounding, Assaults, Stealing/Theft, Rape/Indecent Assault, Kidnapping, False Pretence/Cheating, Arson, Forgery, Breach of Peace and Unnatural offences as the major crimes committed or concentrated offences.

Component 2 identify Forgery, Unlawful possession, Store Breaking and Stealing/Theft as the major crimes committed. It has a negative relationship (decrease) with Forgery and Unlawful possession and a positive relationship (increase) with Store breaking and Stealing/Theft.

Component 3 identify Burglary/Stealing only as the concentrated offence. This implies that, Burglary/Stealing is the daily offence in this location.

Component 4 increases with Suicide and decreases with House breaking. We can viewed this location as measure of new common offence while the other offence is becoming history or reducing.

Component 5 identify Arson and Child Stealing and Component 6 identify no specific crime as prevalence.

Component Matrix ^a									
	Principal Component								
	1	2	3	4	5	6			
MURDER	0.760	0.137	-0.257	-0.066	0.137	0.348			
ARMED ROBBERY	0.438	-0.340	-0.481	0.209	-0.189	0.427			
SUICIDE	0.267	0.455	-0.496	0.532	-0.061	0.154			
GRIVIOUS HARM/WOUNDING	0.652	0.184	0.068	0.231	-0.376	-0.082			
ASSAULTS	0.796	-0.060	0.199	0.061	-0.138	-0.352			
BURGLARY/STEALING	0.111	0.333	0.842	0.178	0.053	0.154			
STORE BREAKING	0.492	0.763	-0.043	0.106	0.195	-0.203			
HOUSE BREAKING	0.469	0.035	-0.050	-0.775	-0.018	0.216			
STEALING/THEFT	0.603	0.583	0.132	0.149	0.273	-0.113			
RAPE/INCEDENT ASSAULT	0.770	0.497	-0.071	0.008	-0.013	0.165			
KIDNAPPING	0.790	-0.422	-0.240	-0.038	0.039	-0.135			
FALSE PRETENCE/CHEATING	0.778	-0.236	0.035	-0.140	-0.191	-0.276			
UNLAWFUL POSSESSION	0.230	0.656	0.313	-0.386	-0.266	0.198			
ARSON	0.531	0.024	-0.332	-0.298	0.598	-0.152			
BREACH OF PEACE	0.719	-0.536	0.114	0.040	-0.193	-0.255			
FORGERY	0.515	-0.532	0.469	0.097	0.090	0.427			
CHILD STEALING	0.181	-0.392	0.281	0.224	0.671	0.071			
UNNATURAL OFFENCE	0.759	-0.300	0.203	0.104	-0.087	0.094			
Extraction Method: Principal Component Analysis.									
a. 6 components extracted.									

The Plot below shows the distribution of the Crimes based on the Components extracted. It take the structure of the cube due to the fact six components were extracted and all the components were represented in the plots. As displayed in the plot above, all the crimes were represented in the Component 1. It implies that, component 1 has a positive relationship with all the crimes associated with the state. We expect increase crime rate in the location represented by component 1. Therefore, Oyo state government and the Police officers need to work hard in order to curtail these menace as urgent as possible since the tendency of increase in Crimes or offences is highly significant.



Component Plot in Rotated Space



V. Conclusion

The findings in this work revealed that, crime rates in Oyo state will drastically increase if care is not taken.

Component1 can be viewed as a measure of how unsafe the location (Oyo state) is in terms of Crime rates because most of the major crimes increases with the component 1.Component 2 can be viewed as a measure of common and uncommon crimes in this location (Oyo state). Forgery and Unlawful possession decreases and Store breaking and Stealing/Theft is increasing.Component 3 suggested Burglary/Stealing as the daily offence in Oyo state. Component 4can be viewed as measure of new common offence while the other offence is becoming history or reducing. Component 5 identify Arson and Child Stealing. Component 6 does not specify any peculiar crime. This components helped us in identifying the distribution of crimes committed in Oyo state.

Thus, identifying the distribution of these crimes in Oyo State will allows the investors, tourist and travelers to measure the level of risk and to plan preventive measures for safeguarding their investments, property and lives.

References:

- [1]. Gulumbe et al (2008); Analysis of Crime Data using Principal Component Analysis.
- [2]. Alexandra Vaughn; Columbia University QMSS 5999 Thesis Spring 2012; Crime rate and concern for time series analysis.
- [3]. International Journal of Academic Research in Business and Social Sciences; The Influence of Sport and Recreation upon Crime Reduction: June 2012, Vol. 2, No. 6 ISSN: 2222-6990
- [4]. Brian Bell and Stephen Machin; Centre for Economic Performance, LSE; The Impact of Migration on Crime and Victimization.
- [5]. Aaron Chalfin University of Cincinnati, Justin McCrary UC Berkeley, NBER; Criminal Deterrence: A Review of the Literature.
- [6]. Michael Shively (June 2005); Study of Literature and Legislation on Hate Crime in America
- [7]. Robin Beaumont; Principal Component Analysis & Factor Analysis:Using SPSS 19 and R (psych package)
- [8]. Gabor, T. (1978). Crime displacement: the literature and strategies for its investigation. Crime and Justice, 6:100 7.
- [9]. Maritz, J. (2010). Honest answers to your questions about investing in Nigeria: Will I have to fear for my safety in Nigeria? June, 14, 2010.http://www.tradeinvest Nigeria. com/news/623915.htm
- [10]. Rencher, A.C. (2002) Methods of Multivariate Analysis. 2nd edition, John Wiley & Son, New York.
- [11]. Richard, A.J. and Dean, W.W. (2001). Applied Multivariate Statistical Analysis. 3rd edition, Prentice-Hall, New Delhi.
- [12]. Cattell, R.B. (1966). The Scree test for the number of factors. Multivariate Behavioral Research, 1,245-276.
- [13]. Durkheim, E. (1933). The division of Labour in Society. Macmillan, New York. Financial (2011). Nigeria crime. Financial Times, 7/11/2011.
- [14]. Jolliffe, I.T. (2002). Principal Component Analysis. 2nd edition, Springer-Verlag, New York. CBN Journal of Applied Statistics Vol. 3 No.2 49
- [15]. Laerd Statistics; How to perform a principal components analysis (PCA) in SPSS
- [16]. Lagos (undated). Crime Rate in Nigeria Free Tips to Deal with Armed Robbery Attacks, Pickpockets, Shoplifting, And 419. http://lagos-nigeria-real-estate-advisor.com/crime-rate.html
- [17]. Louis, S., Cookie, W. S., Louis, A. Z. and Sheldon, R. E. (1981). Human Response to Social Problems. The Dorsey Press, Illinois.