PRINT ISSN 1119-8362
Electronic ISSN 2659-1502

Full-text Available Online at
https://www.ajol.info/index.php/jasem
http://www.bioline.org.br/ja

J. Appl. Sci. Environ. Manage.
Vol. 27 (5) 1003-1008 May 2023

Use of Principal Component Analysis for Evaluation of Causes of Insecurity and Crime Rate Investigation in Niger State, Nigeria

AHMED, SS; USMAN, NM; JIBRIN M; SAIDU, U; AHMED, RT

Department of Mathematics and Statistics, School of Natural and Applied Sciences (SNAS), Niger State Polytechnic, Zungeru, Niger State, Nigeria

**Corresponding Author Email: ahmedsule710@gmail.com; Tel: 08036154396*
Co-Authors Email: usmannuhumariga@gmail.com; elhajee24@gmail.com; saiduumaru3656@gmail.com; rufaitete205@gmail.com

ABSTRACT: In this research, Principal Component Analysis (PCA) is utilized for evaluation of causes of insecurity and crime rate investigation in Niger State, Nigeria. Data obtained based on the Principal Component Analysis reveals 76.74% total variation. The outcome extracted three PCs out of the 13 original variables, which implies a great dimensionality reduction, namely: Population, Population Density and Sex Ratio. The paper suggested that, addressing the root causes of crime such as population, drug/arrest, domestic product and sex-ratio will mitigate the insecurity and as well as crime rate in Niger state, Nigeria.

DOI: **https://dx.doi.org/10.4314/jasem.v27i5.18**

Cite this paper as: AHMED, S. S; USMAN, N. M; JIBRIN M; SAIDU, U; AHMED, R. T. (2023). Use of Principal Component Analysis for Evaluation of Causes of Insecurity and Crime Rate Investigation in Niger State, Nigeria. *J. Appl. Sci. Environ. Manage.* 27 (5) 1003-1008

Dates: Received: 17 February 2023; Revised: 08 April 2023; Accepted: 16 April 2023
Published: 30 May 2023

Keyword: Population Density, Components, Insecurity, Crime Rate, Poverty Rate

Niger state has one of the most alarming crime rates in the country. Between 2019 and 2022, bandit killed a lot of people and possibly most attacks in Mariga, Rafi, Munya and other neighboring LGAs. It was said that, as the case with the rest of the countries, image is merely an exaggeration. And also, added that, Niger's rural areas have more problems with crime than the urban areas. Most crimes are however, purely as a result of poverty. This has led to the formation of various vigilante groups, to combat crimes in some parts of the state. Usman *et al* (2012), in their analysis carried out using NCSS and GESS 2007 Software. The outcome of the results shows that three Principal components have been maintained using the Scree plot and Loading plot giving that correlation exist between crimes against persons and crime against properties. Nigeria record one of the highest crime rates in the world. Killing often accompanies minor burglaries; Influential Nigerians live in high-security apartments. Security men in some states are enforced to "short on sight" (financial times, 2011). No disagreement from both macro and micro side of studies that the rate of crime in Nigeria has reached an unacceptable level. The environment is said to play significant role in determining criminal behaviour. Factors within the environment that mostly influence criminal behaviour include poverty, employment, corruption, urbanization, family, moral decadence, poor education, technology, child abuse, drug trafficking and abuse, architectural or environmental design (Oyebanji, 1982) and (Akpan, 2002) have attribute the current crime problem. Some of these assaults happened as domestic violence, while other are inflicted by criminals on guards specifically under the volatile situation (Oshunkeye, 2014). These assaults have resulted in default joint partial loss of listening and vision, permanent disfigurement, scars from burns, knives and machete wound (Ikeoh, 2012). However, from human and sociological effect of crime, there is a significant

**Corresponding Author Email: ahmedsule710@gmail.com*

economic cost to the country in which rate of crime is high. Such economic consequences include decrease in labor market participation. Reduce productivity (krug *et al* 2012). The increase in urban crime rate in Nigeria is among the major social problems facing the country in recent time. The concentration of social vices in urban centres all over the world, particularly in Niger state is an indication of the breakdown of urban system. Also, many urban centres of this country today, illegal activities are assuming dangerous tendencies as it affect lives and properties. Therefore, decreasing the quality of life of citizen (Agboola, 2010; Ahmed, 2009). In the southwestern part of Nigeria, It was observed that the highest and commonly committed crimes are assault, grievous harm and wounding, theft/stealing, burglary, housebreaking, false presence unlawful arms possession and strand of peace (Femi *et al*, 2015). The situation confirmed from the report of the International Crime Victim Survey (ICVS) shows that, the report was conducted on six major continents for the 1989 – 1996 period that more than half of the urban respondents reported being victimized at least once regardless of what part of the world they inhabit (Ackerman and Murray, 2014). The most significant aspect of criminal justice system is the police. Criminal justice system can be defined as a procedure of processing the person accused of committing crime from arrest to the final disposal of the case. Meanwhile, for the past three decades there have been serious dissatisfaction and public criticisms over the conduct of the police. Now, what are the causes of the police failure in preventing and controlling the crimes? So many reasons can be attributed to the problem. Some factors are inadequate manpower, equipment and professionalism (Danbazau, 2007), the computation of PCA reduced to an eigenvalue - eigenvector problem. It is solved either on a correlation or a covariance matrix. If some group of measure constitutes the score of the numerous variables, the researchers may like to combine the score of the many variables into smaller number of super variable to form the group of the measure (Jolliffe, 2002). In another research carried out by (Ayoola, 2008), lack of integrity, transparency and accountability in the management of public funds, particularly at all the stages of government have been identified as the factors responsible for the native corruption that has eaten deep into the fabric of the Nigerian society over the years. Among the fundamental challenges in crime mapping and analysis is pattern recognition. Efforts and methods to detect crime hot-spots, or geographic areas of elevated criminal activity, are wide ranging. For aggregate data, such as total crime events in a census tract(s), (Grubesic, 2006). This paper explain methods for crime incident projecting by focusing upon geographical areas of concern that transcend traditional policing boundaries (Corcoram, 2003). In a study to identify the statistical relationship between crime and socio-economic status in Ottawa and Saskatoon, the PCA was used to replace a set of variable with a smaller number of components, which are made up of inter-correlated variable representing the original data set as possible (Exp, 2008). Principal component analysis can also be used to examine the overall criminality. When the first eigenvector shows approximately equal loadings on all variables than the first PC measure the overall crime rate. In printcom (2003) for 1997 US crime data, the overall crime rate was examined from the first PC, and the same outcome was achieved (Hurdle *et al* 2007) for the 1985 US crime data. The objectives of this research, Principal Component Analysis (PCA) was utilized for evaluation of causes of insecurity and crime rate investigation in Niger state, Nigeria

MATERIALS AND METHODS

Data Collection: The data contains thirteen (13) variables that fall under several general characteristics categories such as population, economic, social and housing characteristics and strength of police personals. It's worth mentioning that we do not include all the common variables listed above because factors like culture and transportation do also affect the crime of a particular area, which in our case are the Local Governments in Niger State, NIgeria. In addition, compared to the country and the state, some statistics are not measurable and applicable.

We collect data of poverty rate, unemployment rate, average household size, youth illiteracy rates and percentage moderate income household livelihood 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), 2019 Presidential election percentage voters turnout from Independent National Electoral Commission (INEC) and crime record (2011) and police strength from Nigeria Police force. We use crime record from 2011 because this is the only recent years for which we could find complete data. The dataset totally has 25 entries; each entry represents the information of a particular Local Government in Niger State, Nigeria

List of Variables: *POP* => Population; *POD* => Population Density; *SXR* => Sex Ratio; *AHS* =>

Average Household Size; *ILR* => Illiteracy Rate (Youth); *MHI* => Moderate Household Incomes; *PVR* => Poverty Rate;
UER => Unemployment Rate; *VOT* => Voters Turnout; *PLS* => Police Strength; *GDP* => Growth Domestics Product; *DOA* => Drug Offenders Arrested; D*SAI* => Drug Seizure Arrested Index; *Method of Evaluation:* The data used in this study was obtained from the Nigerian Bureau of Statistics as reported by the Nigerian Police on criminal issues in Niger State (2017).

Principal Component Analysis: Principal Component Analysis (PCA) will 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. We can explain 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 idea of principal component transformation is to look for few variables less than p derived variables that preserved most of the information given by the variance of the p random variables.

Let the random vector $X' = X_1, X_2, X_{3,} ---, X_p$ have the covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \lambda_p \geq 0$. Consider the linear combinations:

$$Y_j = \alpha_j' X = \alpha_{j1}X_1 + \alpha_{j2}X_2 + \alpha_{j3}X_3 + \cdots + \alpha_{jp}X_p = \sum_{k=1}^{p} \alpha_{jk}X_p \quad 1$$

Such that $j = 1, 2, \ldots, p$ are the elements of $X \, and \, \alpha_{j1}, \alpha_{j2}, \alpha_{j3}, \ldots, \alpha_{jp}$ are the components of $\alpha_j$vector of p^{th} term.

Then, $Var(Y_j) = \alpha_j' \sum a_j \quad j = 1, 2, \ldots, p \quad 2$
$Cov(Y_J . Y_k) = \alpha_j' \sum a_k \quad j = 1, 2, \ldots, p \quad 3$

Principal Component Procedure: The principal components are those uncorrelated linear combinations $Y_1, Y_2, Y_3, \ldots, Y_p$ whose variances in (1.2) are as large as possible. In finding the PCs we concentrate on the variances. The first step is to look for a linear combination $a'X$ with maximum variance, so that

$$\alpha_1' X = \alpha_{11}X_1 + \alpha_{12}X_2 + \alpha_{13}X_3 + \cdots + \alpha_{1p}X_p = \sum_{k=1}^{p} \alpha_{1k}X_p \quad 4$$

Then, we look for the linear combination $\alpha_2' X$ uncorrelated with $\alpha_1' X$ having maximum variance and so on, hence at the k^{th} stage a linear combination $\alpha_k' X$ is found that has maximum variance subject to being uncorrelated with $\alpha_1' X$, $\alpha_2' X$, $\alpha_3' X$, …, $\alpha_{k-1}' X$. The k^{th} derived variable $\alpha_k' X$ is the k^{th} principal component. Up to p principal components can be found, but we would hope to stop after the q^{th} stage for $(q \leq p)$, i.e. when most of the variation in X would have been accounted for by q^{th} principal components.

Note:
The variance of a principal component is equal to the eigenvalue corresponding to that principal component,
$Var(Y_j) = \alpha_j' \sum a_j = \lambda_j \qquad j = 1, 2, 3, \ldots, p$
The total variance in data set is equal to the total variance of principal components

$$\sigma_{11} + \sigma_{22} + \sigma_{33} + \cdots + \sigma_{pp} = \sum_{j=1}^{p} Var(X_j)$$
$$= \lambda_1 + \lambda_2 + \lambda_3 + \cdots + \lambda_p$$
$$= \sum_{j=1}^{p} Var(Y_j)$$

The data would be 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 $X' = [X_1, X_2, X_3, \ldots, X_p]$ the corresponding standardized variables are $z = \left[Z = \frac{(X_j - \mu_j)}{\sqrt{\sigma_j}}\right]$ for $j = 1, 2, 3, \ldots, p$ in matrix notation, $Z = \left(\theta^{1/2}\right)^{-1}(X - \mu)$, Where $\theta^{1/2}$ is the diagonal standard deviation matrix and it's a unit.

Thus, $E(Z) = 0 \; and \; Cov(Z) = \rho$.

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 Y_j refers to the j^{th} PC and (λ_j, a_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 j^{th} PC of the standard variables $Z' = [z_1, z_2, z_3, \dots, z_p]$ with $cov(Z) = \rho$, is given by

$$Y_j = \alpha_j' Z = a'\left(\theta^{1/2}\right)^{-1}(X - \mu) \qquad 5$$

So that $\sum_{j=1}^{p} Var(X_j) = \sum_{j=1}^{p}(Z) = \rho \qquad for \;\; j = 1, 2, 3, \dots, p$

In this case, (λ_1, α_1), (λ_2, α_2), (λ_3, α_3), …, (λ_p, α_p) are the eigenvalue - eigenvector pairs for ρ with $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \lambda_p \geq 0$.

Interpretation of the Principal Components: The loading or the eigenvector $\alpha_j = \alpha_1, \alpha_2, \alpha_3, \dots \alpha_p$ is the measure of the importance of a measured variable for a given PC. When all elements of α_1 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. The plot of the first two or three loadings against each other enhances visual interpretation.

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

$$Y_{ij} = \alpha_{j1}' X_{i1} + \alpha_{j2}' X_{i2} + \alpha_{j3}' X_{i3} + \cdots + \alpha_{jp}' X_{ip}$$
$$i = 1, 2, 3, \dots, p \qquad j = 1, 2, 3, \dots, p$$

The plot of the first two or three PCs against each other enhances visual interpretation.

The proportion of variance will be used to tell us the PC that best explained the original variables and the measure of how well the first q PCs of Z explain the variation is given by:

$$\phi_q = \frac{\sum_{j=1}^{q} \lambda_j}{p} = \frac{\sum_{j=1}^{q} V(z_j)}{p} \qquad 6$$

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.

RESULTS AND DISCUSSION

The major types of crime in the analysis are: crime against person which includes: Murder, Grievous Harm and Wounding (GHW), assault, kidnapping, rape and crime against property which include armed robbery, theft and house breaking. Table 1: The correlation matrix has displayed different levels of correlation between the variables. Looking at the table, there are very low correlations between populations on one hand and illiteracy rate and poverty rate; also between population density and sex ratio, average household size and poverty rate, etc. But a high correlation between population and unemployment rate, police strength, growth domestic product, drug offenders arrested and drug seizure/arrest index. And the determinant of correlation matrix equal to 0.000. Therefore, multicollinearity is not a problem for this data there is no need to consider eliminating any variable at this stage. Table 2: shows the result of KMO measure of sampling adequacy. The KMO statistic varies between 0 and 1. Kaiser (1970) recommends accepting values greater than 0.5 as acceptable (value below this should lead you to either collect more data or rethink which variable to include). For these data the value is 0.720503 which is greater than 0.5: so, we should be confident that PCA is appropriate for these data.

Table 1: Correlation Matrix

	POP	POD	SXR	AHS	ILR	MHI	PVR	UER	VOT	PLS	GDP	DOA	DSAI
POP	1												
POD	0.240	1											
SXR	0.225	0.079	1										
AHS	0.130	0.086	-0.195	1									
ILR	-0.018	-0.676	-0.152	0.1134	1								
MHI	0.247	0.195	0.206	-0.336	-0.134	1							
PVR	-0.113	-0.068	-0.099	0.312	-0.024	-0.736	1						
UER	0.617	0.181	0.028	0.131	-0.011	-0.048	-0.066	1					
VOT	-0.779	-0.167	-0.063	-0.033	-0.084	-0.040	0.036	-0.571	1				
PLS	1.000	0.239	0.224	0.132	-0.017	0.247	-0.112	0.619	-0.778	1			
GDP	1.000	0.240	0.225	0.130	-0.018	0.247	-0.113	0.617	-0.779	1	1		
DOA	0.998	0.250	0.206	0.145	-0.026	0.215	-0.092	0.632	-0.779	0.998	0.998	1	
DSAI	0.999	0.249	0.227	0.129	-0.033	0.245	-0.113	0.628	-0.777	0.9996	0.999	0.998	1

Determinant = .0000

Table 2: Kaiser-Meyer-Olkin (KMO)

POP	POD	SXR	AHS	ILR	MHI	PVR	UER	VOT	PLS	GDP	DOA	DSAI	
0.705	0.445	0.636	0.560	0.265	0.536	0.418	0.634	0.784	0.863	0.706	0.836	0.927	0.721

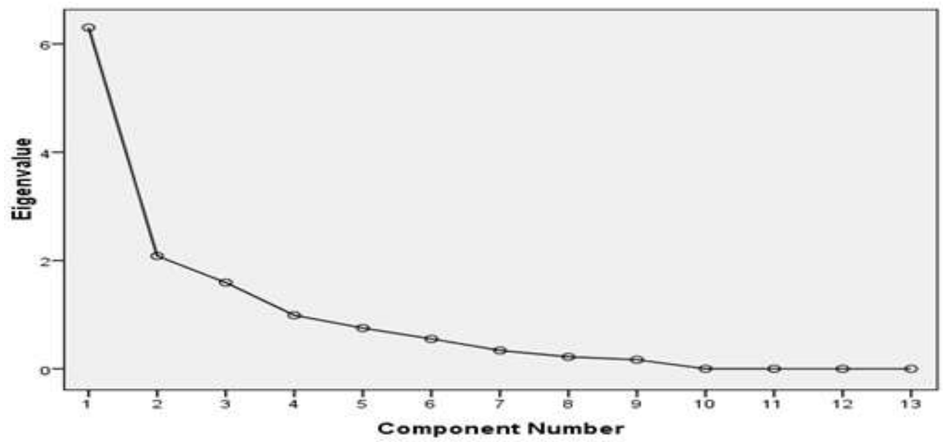

Fig. 1: Scree Plot of the Principal Components

Table 3: Eigenvalues of the Correlation Matrix *Total Variance Explained*

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.300	48.463	48.463	6.300	48.463	48.463	6.169	47.455	47.455
2	2.083	16.024	64.486	2.083	16.024	64.486	2.079	15.991	63.446
3	1.593	12.251	76.737	1.593	12.251	76.737	1.728	13.292	76.737
4	0.989	7.610	84.347						
5	0.753	5.793	90.140						
6	0.553	4.254	94.394						
7	0.339	2.605	96.999						
8	0.221	1.700	98.700						
9	0.167	1.286	99.985						
10	0.002	0.012	99.997						
11	0.000	0.002	100.000						
12	0.000	0.000	100.000						
13	0.000	0.000	100.000						

Extraction Method: Principal Component Analysis.

The first and very important step of PCA is to determine the number of PCs. Here, the Kaiser's criterion is used 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 leaves us with first three PCs that explain up to 76.7% of the total variability of the data set as shown in the Table 3. We can also use the scree plot to choose the number of PCs to retain, as shown in Figure 1 above, the "elbow" appears at the 4th PC, therefore the first three PCs should be retained which account for 76.7% of the total variance, as a valuable reduction in dimensionality. Therefore, it is a fact that the first three PCs accounted for 76.737% of total variance of the original variables and simultaneously reduce the data dimension from 13 to 3.

Table 4: Rotated Component Matrix

	Component		
	1	2	3
POP	0.983	0.103	0.074
POD	0.190	0.042	0.886
SXR	0.175	0.348	0.183
AHS	0.194	-0.635	0.023
ILR	0.066	-0.073	-0.925
MHI	0.141	0.879	0.117
PVR	-0.059	-0.845	0.068
UER	0.702	-0.112	0.049
VOT	-0.840	0.016	0.035
PLS	0.984	0.102	0.073
GDP	0.983	0.103	0.074
DOA	0.985	0.071	0.084
DSAI	0.983	0.102	0.087

The next step is to look at the content of variables that load onto the same factor to try identifying common themes. In Table 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). It can be seen clearly which variables are dominant in each PC. For example in PC2, Moderate Household Income have relative large absolute value, hence PC2 denotes as more economical factor. Similarly, view PC3 as Population Density factor. However, it is kind of difficult to define PC1 in which Population, Unemployment, Police, GDP, Arrest and seizure all have relative large absolute value of greater than 0.7 level, which indicates that interpretation of PCs is one major disadvantage of PCA.

Conclusion: Principal Component Analysis has been successfully applied into our data by extracting three PCs out of the 13 original variables, which implies a great dimensionality reduction. In addition, these PCs account for high percentages of variance of the original dataset, thus suggesting that we do not loss much information. Highly correlated original variables with a particular PC can serve to determine the label of that PC. Moderate Household Income, Unemployment Rate and population density collaboratively make contribution to PC2. PC3 and PC1.

REFERENCES

Ackerman, WV; Murray, AT (2014). "Assessing spatial patterns of crime in Lima Ohio". USA Cities, 21(5): 432-437. Published by Elsevier Ltd. Printed in Great Britain

Agboola, B; Ofoegbu, F (2010). Access to University Education in Nigeria. Published 27 July, 2010 P. 17.

Ahmed, YA (2009). Trend and pattern of urban crime in southwestern Nigeria, Unpublished Ph.D. Thesis, University of Ilorin, Nigeria.

Akpan, AU (2002). Notion of causes of crime among Nigerians. *Inter. J. Soc. Policy Issues:* 1(1): 36-40.

Ayoola, AO (2008). Ayoola identifies causes of corruption. This Day, 20/2/2008.

Corcoran, JJ; Wilson, ID; Ware, JA (2003). Predicting the geo-temporal variations of Crime and disorder. *Inter. J. Forecasting*. 19(4): 623 – 634

Danbazau, AB (2007). Criminology and Criminal Justice. 2nd edition. Ibadan: Spectrum Books Limited.

Exp (2008). http://www.justice.gc.ca/eng/pi/rs/rep-rap/2006/rr06/6/p5.html

Femi, JA; Adeyemi, MA; Jabaru, SO (2015). On the Estimation of Crime Rate in the Southwest of Nigeria: Principal Component Analysis Approach.

Financial Times (2011). London, Thurs. 30/04/2009.

Grubesic, TH (2006). On the Application of Fuzzy Clustering for Crime Hot Spot Detection. *J. Quantitative Criminology*. Springer Netherlands. 22(1): 77-105

Hurdle, W; Zdenek, H (2007). Multivariate statistics: Exercises and solution. New York: Springer-verlag.

Ikeoh, MU; (2012). Criminal Victimization in Nigeria: Pattern and Trend. In Alemike, E.E.O & Chukwuma, I.C (eds) Crime victimization, Safety and Policing in Nigeria Lagos; CLEEN Foundation

Jolliffe, IT; (2002). Principal Component Analysis 2nd ed. New York: Springer- Verlag.

Krug, E; Dahlberg, LL; Mercy, JA *et al* (2012). Access Continuing Education. *https://www.accesscontinuingeducation*

Oshunkeye, S (2014). Africa journalism in culture of brown envelopes. A paper presentation at the Media Project's Regional Conference in Accra, Ghana. Retrieved; August 25, 2018 from *J. vs brown envelops.*

Oyebanji, JO; (1982) Economic Development and the Geography of crime: an Empirical Analysis. GeoJournal. 6(5):

Printcom (2003): http://suport.sas.com/onlinedoc/912/getDoc

Richard, AJ; Dean, WW (2001). Applied Multivariate Statistical Analysis, 3rd ed. New Delhi: Prentice-Hall.

Usman, U; Yakubu, M; Bello AZ (2012). An investigation on the rate of crime in Sokoto State using Principal Component Analysis. *Nig. J. Basic and Appl. Sci.* 20(2): 152-160.