Analysis of Nigeria's Crime Data: A Principal Component Approach using correlation matrix

Atanu, E. Y.

Department of Statistics, Federal Polytechnic of Oil and Gas, P.M.B. 5027, Bonny Island, Rivers State, Nigeria.

DOI: 10.29322/IJSRP.9.01.2019.p8503
http://dx.doi.org/10.29322/IJSRP.9.01.2019.p8503

Abstract

This paper analyses Nigeria's crime data consisting of eight major crimes reported to the police in 2017. The crimes consist of robbery, theft, house breakings, grievous harm and wounding, murder, rape, and assault. Principal component analysis (PCA) was employed to explain the correlation between the crimes. The result shows a significant correlation between armed robbery, theft and grievous harm and wounding. While Kebbi state has the lowest crime rate, Lagos state has the highest overall crime rate in the state. Theft, house breakings, grievous hurt and wounding, murder, rape is 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 and giving the correlation between them the study concludes that some of the crime data can be used to predict each other and an interesting fact that kidnapping is only positively correlated with murder.

Keywords: Crime rate, Principal component analysis, Crimes in Nigeria, Crime data, Kidnapping.

1.0 Introduction

Crime as we know has no universal definition. It can be linked to changes in social, political, psychological and economic conditions. An act may be a crime in one society, but not in another [1]. Activities like prostitution, adultery and homosexuality between consenting adults have been wholly or partially removed from state criminal law in America [2] but are considered as crimes in place such as Saudi Arabia and Nigeria etc. The constant changes in time also change the perception of society on crime. Today, Child marriage is a crime in most countries which was not always so and in some countries right now polluting the air or water can be considered a crime. Pollution causes few problems and receives little attention in colonial days [3]. Therefore, the perception of an "act" to be a crime varies with time and space. In addition, many scholars have defined crime in different views, mostly bordering on ethical and ideological orientation.

One of the fundamental techniques to combat criminal activities is the better understanding of the dynamics of crime. Crime is often thought of as a moral threat and injurious to the society. It afflicts the personality of individual and his property and lessens trust among members of the society [4]. The causes of crime are multiple and could be traced to bio-genetic factors, such as genetic mutation and heredity, psychological factors, such as personality disorders and sociological factors, such as learning and environment [5]. The diverse differences in geographical areas in terms of population density, demographic characteristics, natural vegetation, location and socio-economic characteristics has rendered crime rate unevenly distributed globally. However, it has been observed that the entire world is experiencing high criminal rate.

Over the years the rate of crime in Nigeria has been on the increase and these crimes are being carried out with perfection and sophistication. This has led to the formation of various vigilante groups, to combat crimes in some parts of the country [6].

Crime is one of the continuous problems that bedevil the existence of mankind. Since the early days, crime had been a disturbing threat to mankind's' personality, property and lawful authority [4]. Today, in the modern complex world, the situation is highly disturbing. Crime started in the primitive days as a simple and less organized issue, and ended today as very complex and organized. Therefore, the existence of crime and its problems have spanned the history of mankind.

Nigeria has one of the most alarming crime rates in the world [7]. In April of 2018, armed robbers killed at least 30 people and possibly more in attacks on a bank and police station in Kwara state, Nigeria [8]. It was [9] who said that as is the case with the rest of the world, image is merely an exaggeration. And also, added that, Nigeria's metropolitan areas have more problems with crime than the rural areas. Most crimes are however, purely as a result of poverty.

Despite the fact that, crime is inevitable in a society [10], various controlling and preventive measures had been taken, and are still being taken to reduce the menace with the introduction of modern scientific and technical methods in crime prevention and control proving to be effective.

In the 1980s, serious crime grew to nearly epidemic proportions, particularly in Lagos and other urbanized areas characterized by rapid growth and change, stark economic inequality and deprivation, social disorganization, inadequate government service and lower law enforcement capabilities which subsequently lead the government to deploy but the military and Police with orders in some states to "shoot on sight" violent criminals [11].

Methodology

Data collection

The data used in this study was obtained from the Nigerian bureau of statistics as reported by the Nigerian Police for 2017.

Principal Component Analysis

Principal component analysis 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 [12].

Let the random vector $X' = X_1, X_2, X_3, \dots, X_P$ have the covariance matrix Σ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p \geq 0$. consider the linear combinations:

$$Y_j = a'_j X = a_{j1}X_1 + a_{j2}X_2 + a_{j3}X_3 + \cdots + a_{jp}X_p = \sum_{k=1}^{p} a_{jk}X_k \quad (1.1)$$

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

Then, $Var(Y_j) = a'_j \sum a_j \qquad j = 1,2, \dots, p \quad (1.2)$

$$Cov(Y_j.Y_k) = a'_j \Sigma a_k \qquad j = 1,2, \dots, p \quad (1.3)$$

Principal Component Procedure

The principal components are those uncorrelated linear combinations $Y_1, Y_2, Y_3, \dots, Y_p$ whose variances in (1.2) are as large as possible [12]. In finding the PCs we concentrate on the variances. The first step is to look for a linear combination $a'\boldsymbol{X}$ with maximum variance, so that

$$a_1'X = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \cdots + a_{1p}X_p = \sum_{k=1}^{p} a_{1k}X_k \quad (1.4)$$

Then, we look for the linear combination $a'_2\boldsymbol{X}$ uncorrelated with $a'_1\boldsymbol{X}$ having maximum variance and so on, hence at the k^{th} stage a linear combination $a'_k\boldsymbol{X}$ is found that has maximum variance subject to being uncorrelated with $a'_1\boldsymbol{X}, a'_2\boldsymbol{X}, a'_3\boldsymbol{X}, \dots, a'_{k-1}\boldsymbol{X}$. The k^{th}

derived variable $a'_k\boldsymbol{X}$ 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 PCs.

Note:

- The variance of a principal component is equal to the eigenvalue corresponding to that principal component,

$$Var(Y_j) = a'_j \sum a_j = \lambda_j \qquad j = 1,2,3,\dots,p$$

- The total variance in data set is equal to the total variance of principal components $\sigma_{11} + \sigma_{22} + \sigma_{33} + \cdots + \sigma_{pp} = \sum_{j=1}^{p} Var(X_j) = \lambda_1 + \lambda_2 + \lambda_3 + \cdots + \lambda_p = \sum_{j=1}^{p} Var(Y_j)$

From [13], 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 $X' = [X_1, X_{2,} X_3, \dots, X_p]$ the corresponding standardized variables are $Z = \left[z_j = \frac{(X_j - \mu_j)}{\sqrt{\sigma_{jj}}}\right]$ for $j = 1,2,3,\dots,p$ 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

$$Y_j = a'_j Z = a'_j(\theta^{1/2})^{-1}(X - \mu) \qquad (1.5)$$

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

In this case, $(\lambda_1, a_1), (\lambda_2, a_2), (\lambda_3, a_3), \dots, (\lambda_p, a_p)$ are the eigenvalue - eigenvector pairs for ρ with $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_p \geq 0$.

Interpretation of the Principal Components:

The loading or the eigenvector $a_j = a_1, a_2, a_3, \dots, a_p$, is the measure of the importance of a measured variable for a given PC. When all elements of a_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 [14]. The plot of the first two or three loadings against each other enhances visual interpretation [15].

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

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

The plot of the first two or three PCs against each other enhances visual interpretation [15].

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:

$$\phi_q = \frac{\sum_{j=1}^{q} \lambda_j}{p} = \frac{\sum_{j=1}^{q} Var(Z_j)}{p} \qquad (1.6)$$

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 [16].

Results and Analysis

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: Correlation of Crime Types (per 10,000 population)

	Armed Robbery	Theft	Kidnapping	Rape	House Breaking	Murder	Grievous harm & wounding	Assault
Armed Robbery	1.000	.689	-.129	.535	.612	.523	.702	.632
Theft		1.000	-.109	.866	.960	.506	.992	.981
Kidnapping			1.000	-.111	-.149	.347	-.1 27	-.182
Rape				1.000	.869	.480	.862	.854
House Breaking					1.000	.469	.948	.954
Murder						1.000	.533	.460
Grievious harm wounding							1.000	.970
Assault								1.000

The correlation matrix above displayed different levels of correlation between the crimes. There is strong positive relationship between armed robbery, theft, rape, house breaking-in, murder and grievous harm and wound (GHW). The relationships were also significant at *p*=.05 significance level, which means that their variables can be used to predict (explain) one another. Similarly, it was observed that the correlations between kidnapping and armed robbery, theft, rape, house breaking-in, grievous harm & wounding (GHW) and assault were negative. Kidnapping was only positively correlated with Murder.

Table 2: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.789
Bartlett's Test of Sphericity	Approx. Chi-Square	332.558
	Df	28
	Sig.	.000

The null hypothesis that the correlation matrix is an identity matrix was rejected at 5% level of significance (Bartlett's test of Sphericity; χ^2 = 332.558, p-value =.000), this implies that the correlation in the dataset are appropriate for factor analysis.
Also, the Kaiser-Meyer-Olkin statistic = 0.789 revealed that adequate sampling is being used for this analysis.

Table 3: Eigen values

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.523	69.032	69.032	5.523	69.032	69.032
2	1.291	16.134	85.166	1.291	16.134	85.166
3	.605	7.566	92.732	.605	7.566	92.732
4	.325	4.059	96.792	.325	4.059	96.792
5	.175	2.188	98.980	.175	2.188	98.980

6	.054	.675	99.655	.054	.675	99.655
7	.023	.293	99.949	.023	.293	99.949
8	.004	.051	100.000	.004	.051	100.000

Extraction Method: Principal Component Analysis.

The eigenvalues, proportion and the cumulative proportions of the explained variance are displayed in Table 3. Considering the eigenvalues and the Scree plot below in figure 1 and according to [17], the rule of thumb is usually applied to retain those factors whose eigenvalues are greater than one. Hence, it would be reasonable to retain the first two PCs i.e. Armed robbery and Theft. So the first 2 PCs can be retained to explain up to 85.166 percent of the total variability.

Figure 1: Scree plot

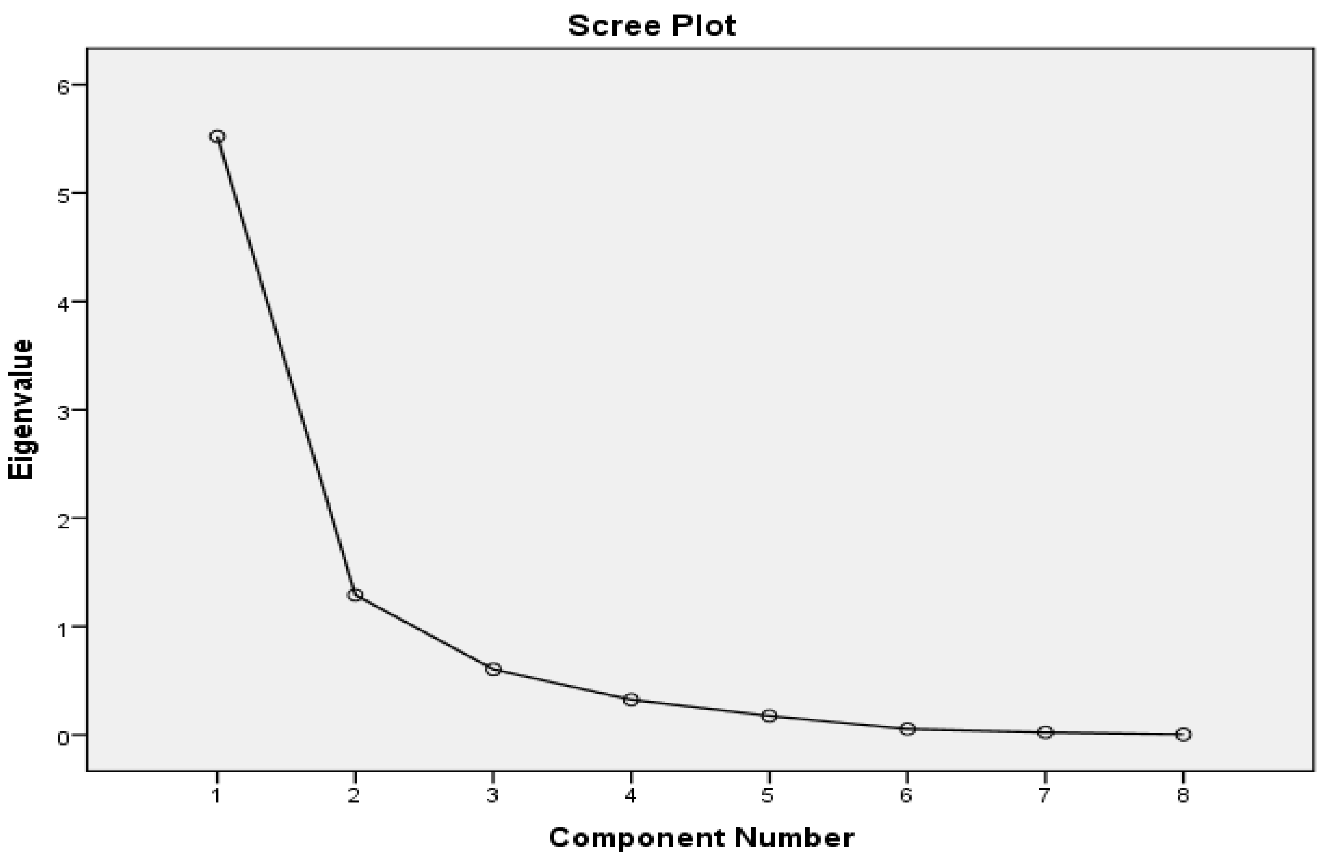


Table 4: Eigen vector

	Component	
	1	2
Armed Robbery	.746	.050
Theft	.984	-.041
Kidnapping	-.119	.918
Rape	.897	-.040
House Breaking	.959	-.089
Murder	.596	.647

Grievious harm wounding	.985	-.038
Assault	.966	-.118

Extraction Method: Principal Component Analysis.

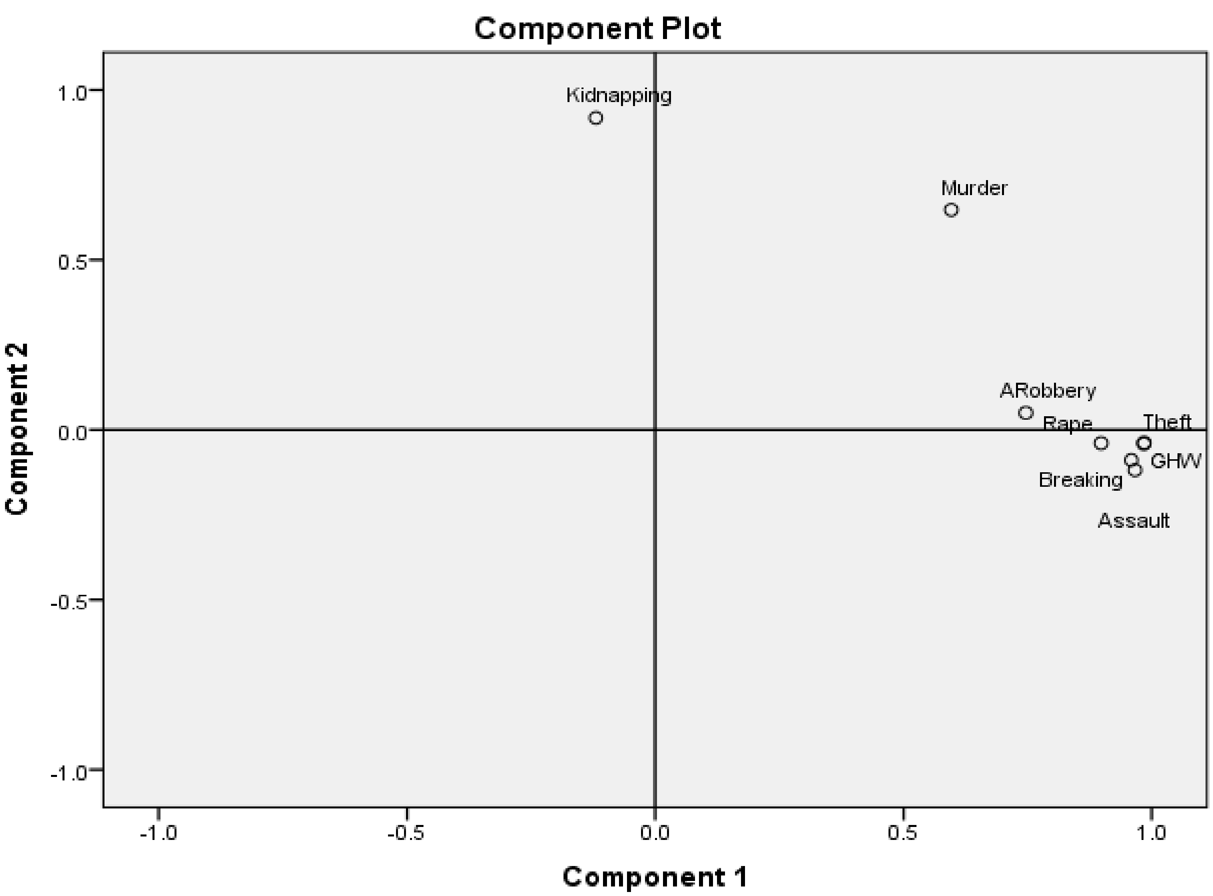
a. 8 components extracted.

From table 4 above, we retained two PCs that explain 85.166 per cent of the total variability of the data set.

Component One: has both positive and negative relationship with all the crimes recorded but majorly it identified; Armed robbery, Theft, Rape, House breaking-in, Grievous harm wounding and Assault.

Component Two: Also has a positive relationship with majorly kidnapping and murder. It has a negative relationship (decrease) with House breaking-in and Assault.

Figure 2: Loading plot of Components



The loading plot in figure 2 above shows the concentration of the selected crimes with theft, assault, etc. been the most committed offence while kidnapping and murder been the less occurring crimes.

CONCLUSION

In this study, Principal Component Analysis was applied to explore the number of principal components to be retained on the eight variables obtained from the criminal data as obtained from Nigeria bureau of statistics which was reported by the Nigeria Police Force in 2017. The results of the statistical analysis proved that two components explain up to 85.17% (Table 3) of the total variability of the data set. The following are the conclusions deduced from the analysis. There is strong positive relationship between armed robbery and rape, grievous hurt and wound (GHW), theft, assault and murder, the relationship were also significant, which means that their variables can be used to predict (explain) one another at p=.05 and at the p-value Kidnapping can be used to predict murder. It was also observed that the correlations in between robbery and burglary, breach of public peace and broken store were negative and insignificant. The state with the highest crime rate is Lagos, while Kebbi state has the lowest crime rate in Nigeria according results obtain from the eight types of crime examined. Two PCs (Armed robbery and Theft) that explains about 85.17 percent of the total variability of the data set are suggested to be retained. The score plot has classified the crimes into two, namely, One: concentrated offences: Armed robbery, Theft, Rape, House breaking-in, Grievous harm wounding and Assault. Two: less concentrated: Kidnapping and Murder. The study also found that Abia state in the south eastern part of Nigeria had the highest number kidnapping cases and that there was no correlation between kidnapping and the selected crime types except murder which from observation was as a result of those perpetrating the crime of kidnapping not intending to harm, rape, etc. but generally, are after ransoms but if ransom is not paid can result in the murder of victims.

REFERENCE:

1. Danbazau, A.B. (2007). Criminology Justice. 2nd edition. Ibadan: Spectrum Books Limited.
2. Feldman, M.P. (1997). The psychological of crime. Cambridge: University Press.
3. WBE (1963). The World Book Encyclopaedia. Vol.4. USA. Field Eterprises Educational Corporation. Pp.908 -912
4. Lois, S., Cookie, W.S., Louis, A.Z. and Sheldon, R.E. (1981). Human Response to Social Problems. Illinois: The Dorsey Press.
5. Sutherland, E.H. (1939). The white-colar criminal. American Sociological Review, 5: 1-12
6. Fajemirokun, F., Adewale, O., Idowu, T., Oyewusi, A., and Maiyegun, B. (2006). A GIS Aproach to crime Mapping and management in Nigeria: A case study of Victoria Island Lagos. www.Oicrf.org
7. Uche, O. (2008). Nigeria Prison Robbed by Criminals. http://www.whichwayNigeria.net /Nigerian-prison-robbed-criminals/
8. “30-killed-in-offa-bank-robbery in Kwara State”. April 6, 2018, Accessed: September 17, 2018 https://www.channelstv.com/2018/04/06/update-30-killed-in-offa-bank-robbery/
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. Durkheim, E. (1933). The division of Labour in Society. Macmillan, New York.
11. Financial Times (2009). London, Thursday 30/04/2009.
12. Jolliffe, I.T. (2002). Principal Component Analysis 2nd edn. New York: Springer-verlag.
13. Shehu U. Gulumbe, H.G. Dikko, and Yusuf Bello. CBN Journal of Applied Statistics Vol. 3 No.2
14. Richard, A.J. and Dean, W.W. (2001). Applied Multivariate Statistical Analysis. 3rd edn, Prentice-Hall, New Delhi.
15. Rencher, A.C. (2002) Methods of Multivariate Analysis. 2nd edn, John Wiley & Son, New York.
16. Soren, H. (2006). Example of multivariate analysis in R – Principal component analysis (PCA). httpgene://tics.agrsci.dk/statistics/courses/Rcourse-Djf2006 / day3/ PCA-notes.
17. Cattell, R.B. (1966). The Scree test for the number of factors. Multivariate Behavioral Research, 1,245-276.
18. Kresta, J.V., Marlin, T.E. and MacGregor, J.F. (1994). Development of inferential process models using PLS. Computer and Chemical Engineering 18(7): 597-611.