Application of Principal Component Analysis to Crime Data, Case Study: Mathare Slums, Nairobi County in Kenya

Wafula Mike Erick[1, *], Samson Wangila Wanyonyi[2], Chris Muchwanju[3]

[1]Department of Mathematics and Actuarial Science, Kenyatta University (KU), Nairobi, Kenya

[2]Department of Mathematics, University of Eldoret, Eldoret, Kenya

[3]Department of Statistics and Actuarial Science, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya

Email address:
okhwafula@gmail.com (W. M. Erick), chrismuchwanju@gmail.com (C. Muchwanju), samsonwangila@gmail.com (S. W. Wanyonyi)
[*]Corresponding author

To cite this article:
Wafula Mike Erick, Samson Wangila Wanyonyi, Chris Muchwanju. Application of Principal Component Analysis to Crime Data, Case Study: Mathare Slums, Nairobi County in Kenya. *American Journal of Theoretical and Applied Statistics*. Vol. 8, No. 1, 2019, pp. 7-17. doi: 10.11648/j.ajtas.20190801.12

Abstract: The study focuses on statistical analysis of causes of crimes in Mathare slums, Nairobi county using data collected via questionnaires in April 2018.The Correlation analysis was carried out to explain the association between the causes of crimes while the principal component analysis (PCA) was used to reduce the dimensionality of the data sets. The Correlation analysis indicates a fairly strong positive relationship between unemployment and drugs and substance abuse which means that their variables can be used to predict one another. PCA analysis reveals that three PCs (drugs and substance abuse, unemployment and neglect from parents) that explains about 52.6% of the total variability of the causes of crimes against person are suggested to be retained. Similarly, two PCs (drugs and substance abuse and unemployment) that explain about 42.2% of the total variability of the causes of crimes against property are suggested to be retained. Generally, the causes of crimes against person and property in Mathare slums are not unique.

Keywords: Crime, Principal Component Analysis, Total Variation, Scree Plot

1. Introduction

Crime can be defined as an unlawful act punishable by state or any other authority.

1.1. Related Work on Crime in Kenya

The existence of crime is as old as mankind and man has always made effort to combat it [1]. The effects of crime across the globe is a serious concern. Kenya being a victim, it has experienced business failures and other progresses particularly in the slum areas due to high rates of criminal activities.

In 2002, UN-Habitat [2] carried out an over four-month victimization survey in Nairobi. The most common causes of crime that were identified by respondents included unemployment and poverty, although general idleness and the quick rewards that crime brings were also noted. A very small minority mentioned the increase in foreigners as the major cause of crime. More details of the report are found in Aki's work [2].

In 2007, Masese [3] did some work on crime and violence trends in Nairobi, Kenya. His study examined the phenomenon of youth crime in Nairobi especially in relation to youth gangs. The study pays special attention to the Mungiki movement and street families as well as examines some of the organized responses to crime of this nature.

In 2010 [4], the Government of Kenya, through the National Steering Committee on Peace building and Conflict Management, Ministry of State for Provincial Administration and Internal Security and with support from UNDP Kenya carried out a three-year peace programme research on the overview of Crime incidents in Nairobi. The results revealed that generally the police were trying their level best to address the problem of crime in Nairobi.

However, several issues were hindering their effective policing. These included inadequate number of police

officers, inadequate equipment and tooling to enable police respond to distress calls, antagonistic relationship between the police and members of the public as well as between the police and other security agencies, Inadequate infrastructure, the issue of limited number of alternative reporting facilities other than police stations and posts, fear of victimization of the members of the public by the criminals and police; inadequate understanding of the crime dynamics in an area among others.

In 2011, Ndikaru [5], studied the types and causes of crime between the four slum areas in the City of Nairobi, and also crime victimization within different demographic characteristics of Nairobi's slum residents. The study was conducted in four major slum areas in the city of Nairobi, namely Kibera, Mukuru, Mathare and Korogocho. Secondary data sources from KNBS, formerly known as CBS, were availed for sampling purposes. A total of 660 respondents were sampled. The study found that most common crimes in the slum areas constituted those that attracted short to medium term punitive measures, according to the law, and over four in ten slum dwellers claimed that they had been victims of crime the previous year. The most effective method of curbing crime was undoubtedly police patrols and community policing initiatives.

Darkey and Kariuki [6] did a study on quality of life in Mathare, Nairobi Kenya. The research revealed that Mathare residents prioritize sanitation, waste management and access to water, electricity, education and healthcare as the most essential services for adding quality to their lives. However, employment opportunities and freedom from fear were also suggested to be equally important.

Andvig and Barasa [7] carried out a study on a political economy of slum spaces, case study: Mathare valley. The paper begins by outlining the general aspects of slum control such as ruling from within and outside, the description of Mathare valley area, countrywide history shaping Mathare and the ruling of internal slum space.

In 2015, Muchwanju et al [8] carried out a study on crime rate in Kenya using a mixed effect regression model based on secondary data obtained from KNBS. The findings from their study suggested that poverty rate, unemployment rate, probability of arrest, population density and police arrest are correlated to all typologies of crime rate considered. The results further suggested that crime rate is better explained at provincial level as compared to country level. They recommended that studies focused on finding key determinants of crime should be done at lower levels of administration such as villages and wards in order to curb the menace. This formed the basis of our study. More recently related studies, include work by Wanjiru and Matsubara [9] as well as work by Mburu [10].

1.2. Study Objectives

The objective of the study is to carry out a principal component analysis of crime in Mathare slums, Nairobi County.

The specific objectives of this study are:

1. To determine the bivariate associations between the causes of crimes against person and property.
2. To carry out the principal component analysis and determine the significant components causing criminal activities in Mathare slums.

1.3. Rationale for Using PCA Analysis

The Principal Component Analysis (PCA) is a data analysis tool usually used to reduce the dimensionality (number of variables) of a large number of interrelated variables, while retaining as much information (variation) as possible [11]. It's used for transforming a set of related (correlated) variables into a set of uncorrelated variables that account for decreasing proportions of the variation of the original observation [12]. Past literature has shown that PCA is very useful in crime analysis because of its robustness in data reduction and in determining the overall criminality in a given geographical area. For instance, Kendall and Ralph [13] classified a city as safe or unsafe in the US Cities by using multivariate methods of principal components, factor analysis, and discriminant analysis to reduce the 14 distinct variables that can affect the crime rate of a city to 6 and 7 important variables that showed a high correlation with all the variables. Principal component analysis did decrease the number of variables to 6 and accounted for 86% of the total variance, while factor analysis decreased the number to 7 and accounted for 79.7% of the total variance [14]. Olufolabo et al [15] analyzed Oyo State crime data which consisted of 8 major crimes reported to the police between the periods of 1996 – 2014. They employed correlation analysis and principal component analysis to explain the correlation between the crimes in the state. Soren [16], in his research employed the method of face-to-face personal interview using a stratified multistage random selection procedure. He applied PCA to analyze the spatial pattern of criminal victimization in the 11 Local Government Areas in Kastina Senatorial Zone.

Rencher [12], carried out Correlation analysis and PCA analysis using Katsina state data which consisted of the average of eight major crimes reported to the police for the period 2006 – 2008. The results showed a significant correlation between robbery, theft and vehicle theft. The PCA suggested retaining four components that explained about 78.94% of the total variability of the data set.

2. Methodology and Procedures

2.1. Data Analysis Procedure

The data that was used in this study was obtained through sampled surveys conducted in six wards within Mathare slums from April 2nd to April 25th 2018. In this study, a total of 235 residents aged 18 years and above were considered for the analysis. The sample size is calculated in terms of the number of estimated adults aged 18 years and above that were in the IEBC register by the end of 2017.

The summary of sample data is displayed in Table 1 where

Ni is the number of residents aged 18 years and above in the ith ward, N is the number of residents aged 18 years and above within Mathare slums (population target) i.e $\sum_i Ni = N$ and ni is the sample size calculated in the ith ward.

Table 1. Number of residents aged 18 years and above sampled from each of the six wards within Mathare slums in April 2018.

	Ward	Population size (N_i)	Sample size, ($ni = \frac{Ni}{N}n$)
1.	Kiamaiko	18,559	38
2.	Mlango kubwa	21,612	44
3.	Hospital	11,208	23
4.	Huruma	22,632	46
5.	Mabatini	18,855	38
6.	Ngei	22,693	46
	Total (N)	115,559	235

The data collected was studied in three ways namely:

    i. Descriptive statistics
    ii. Correlations analysis and
    iii. Principal Component Analysis

2.2. Principal Component Procedure

Suppose that we have a random vector X

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}$$

With population variance-covariance matrix

$$Var(X) = \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1P} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2P} \\ \vdots & & \ddots & \vdots \\ \sigma_{P1} & \sigma_{P2} & \cdots & \sigma_P^2 \end{pmatrix}$$

Consider the linear combinations:

$$Y_1 = e_{11}X_1 + e_{12}X_2 + \ldots + e_{1P}X_P$$
$$Y_2 = e_{21}X_1 + e_{22}X_2 + \ldots + e_{2P}X_P$$
$$\vdots$$
$$Y_P = e_{P1}X_1 + e_{P2}X_2 + \ldots + e_{PP}X_P$$

Each of these can be thought of as a linear regression, predicting Y_i from $X_1, X_2, \ldots, X_P$. There's no intercept, but $e_{i1}, e_{i2}, \ldots, e_{ip}$ can be viewed as regression coefficients.

Note that Y_i is a function of the random data, and so is also random. Therefore it has a population variance

$$Var(Y_i) = \sum_{k=1}^{p} \sum_{l=1}^{p} e_{ik} e_{il} \sigma_{kl} = e_i' \Sigma e_i$$

In addition, Y_i and Y_j have population covariance:

$$Cov(Y_i, Y_j) = \sum_{k=1}^{p} \sum_{l=1}^{p} e_{ik} e_{jl} \sigma_{kl} = e_i' \Sigma e_j$$

Collect the coefficients e_{ij} into the vector

$$e_i = \begin{pmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{ip} \end{pmatrix}$$

First Principal Component (PCA 1): Y_1

This is the linear combination of x-variables that has maximum variance (among all linear combinations). It accounts for as much variation in the data as possible.

Specifically, we define coefficients $e_{11}, e_{12} \ldots e_{1p}$ for the first component in such a way that its variance is maximized, subject to the constraint that the sum of the squared coefficients is equal to one. This constraint is required so that a unique answer may be obtained.

More formally, select $e_{11}, e_{12} \ldots e_{1p}$ that maximizes

$$Var(Y_1) = \sum_{k=1}^{p} \sum_{l=1}^{p} e_{1k} e_{1l} \sigma_{kl} = e_1' \Sigma e_1$$

Subject to the constraint that

$$e_1' e_1 = \sum_{j=1}^{p} e_{1j}^2 = 1$$

Second Principal Component (PCA 2): Y_2

This is the linear combination of x-variables that accounts for as much of the remaining variation as possible, with the constraint that the correlation between the first and second component is 0

Select $e_{21}, e_{22}, \ldots, e_{2p}$ that maximizes the variance of this new component.

$$Var(Y_2) = \sum_{k=1}^{p} \sum_{l=1}^{p} e_{2k} e_{2l} \sigma_{kl} = e_2' \Sigma e_2$$

Subject to the constraint that the sums of squared coefficients add up to one,

$$e_2' e_2 = \sum_{j=1}^{p} e_{2j}^2 = 1$$

along with the additional constraint that these two components are uncorrelated,

$$Cov(Y_1, Y_2) = \sum_{k=1}^{p} \sum_{l=1}^{p} e_{1k} e_{2l} \sigma_{kl} = e_1' \Sigma e_2 = 0$$

All subsequent principal components have this same property- they are linear combinations that account for as

much of the remaining variation as possible and they are not correlated with the other principal components.

We will do this in the same way with each additional component. For instance:

i^{th} Principal Component (PCA i): Y_i

We select $e_{i1}, e_{i2}, \ldots \ldots, e_{ip}$ to maximize;

$$Var(Y_i) = \sum_{k=1}^{p} \sum_{i=1}^{p} e_{ik} e_{il} \sigma_{kl} = e_i' \Sigma e_i$$

Subject to the constraint that the sums of squared coefficients add up to one...along with the additional constraint that this new component is uncorrelated with all the previously defined components.

$$e_i' e_i = \sum_{j=1}^{p} e_{ij}^2 = 1$$

$$Cov(Y_1, Y_i) = \sum_{k=1}^{p} \sum_{l=1}^{p} e_{1k} e_{il} \sigma_{kl} = e_1' \Sigma e_i = 0$$

$$Cov(Y_2, Y_i) = \sum_{k=1}^{p} \sum_{l=1}^{p} e_{2k} e_{il} \sigma_{kl} = e_2' \Sigma e_i = 0$$

$$.$$
$$.$$
$$.$$

$$Cov(Y_{i-1}, Y_i) = \sum_{k=1}^{p} \sum_{l=1}^{p} e_{i-1k} e_{il} \sigma_{kl} = e_{i-1}' \Sigma e_i = 0$$

$$Cov(Y_1, Y_2) = \sum_{k=1}^{p} \sum_{l=1}^{p} e_{1k} e_{2l} \sigma_{kl} = e_1' \Sigma e_2 = 0$$

Therefore, all principal components are uncorrelated with one another.

3. Results and Analysis

3.1. Causes of Crimes Committed Against Persons

Crimes committed against person includes crimes such as murder or attempt to murder, rape, assault, wounding, riots, molestation, kidnapping and abduction.

Table 2. Descriptive statistics of causes of crimes against persons.

	Cause of crime	Rating (scale: 5-strongly agree, 4-agree, 3-neutral, 2-disagree and 1-strongly disagree)					Number of respondents	mean	Standard deviation
		Strongly agree	agree	Neutral	disagree	Strongly disagree			
1.	Drugs and substance abuse	149 (63.4%)	54 (23.0%)	11 (4.7%)	15 (6.4%)	6 (2.6%)	235	4.383	1.012
2.	Unemployment	147 (62.6%)	47 (20.0%)	20 (8.5%)	15 (6.4%)	6 (2.6%)	235	4.336	1.043
3.	Neglect from parents	49 (20.9%)	58 (24.7%)	86 (36.6%)	32 (13.6%)	10 (4.3%)	235	3.443	1.094
4.	School drop outs	92 (39.1%)	63 (26.8%)	56 (23.8%)	20 (8.5%)	4 (1.7%)	235	3.932	1.060
5.	Poor role models	74 (31.5%)	63 (26.8%)	60 (25.5%)	27 (11.5%)	11 (4.7%)	235	3.689	1.166
6.	Political or racial differences	50 (21.3%)	57 (24.3%)	75 (31.9%)	40 (17.0%)	13 (5.5%)	235	3.387	1.158
7.	Police siding with criminals	72 (30.6%)	53 (22.6%)	59 (25.1%)	34 (14.5%)	17 (7.2%)	235	3.549	1.261
8.	poor infrastructure within r. estates	84 (35.7%)	55 (23.4%)	49 (20.9%)	35 (14.9%)	12 (5.1%)	235	3.698	1.240
9.	lack of natural surveillance	63 (26.8%)	67 (28.5%)	69 (29.4%)	22 (9.4%)	14 (6.0%)	235	3.609	1.151
10.	poor community cohesion	85 (36.2%)	64 (27.2%)	48 (20.4%)	27 (11.5%)	11 (4.7%)	235	3.787	1.186

Looking at Table 2, one can conclude that:

1. On average drugs and substance abuse, unemployment and school drop outs are the most important variables that influences crimes against persons in Mathare slums. They have the highest mean scores at 4.383, 4.336 and 3.932 respectively. On the other hand,

neglect from parents and political or racial differences have the least influence on crimes against persons with mean scores of about 3.443 and 3.387 respectively.

2. The number of respondents who agree and strongly agree that drugs and substance abuse is a cause of crimes against persons is 86.4%. This is followed closely by unemployment and school drop outs at 82.4% and 65.9% respectively.

3. Neglect from parents as a variable has the highest number of undecided respondents of whether it's a cause of crimes against persons in Mathare slums. It has at 36.6% of undecided respondents. This is followed by political or racial differences at 31.9%.

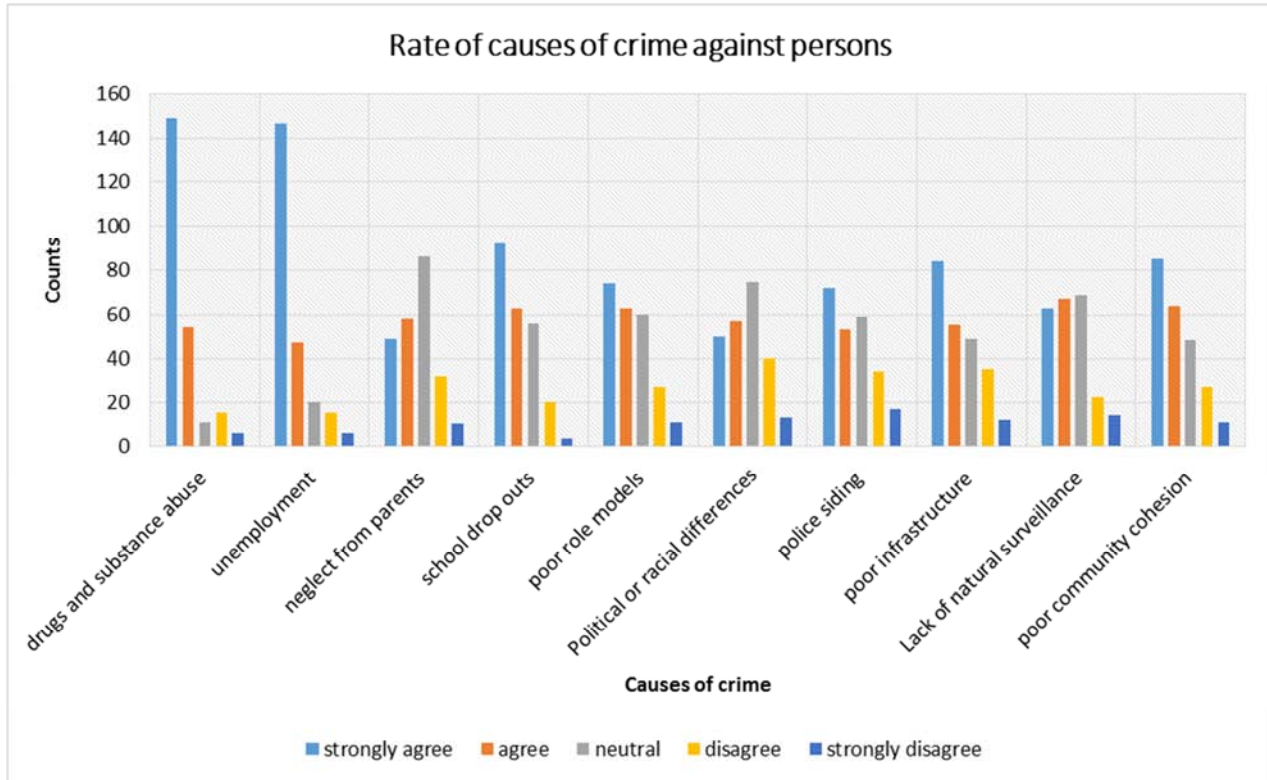A representation of results in table 2 is as shown in figure 1 below.

Figure 1. *Bar graph showing the rate of causes of crimes against persons in Mathare slums.*

Table 3. *Correlation matrix for causes of crime against persons in Mathare slums.*

		Drugs	Unemployment	Neglect from parents	School drop outs	Poor role models
Drugs	Pearson	1	.449**	.136*	.200**	.246**
	Sig(2-tailed)		.000	.037	.002	.000
Unemployment	Pearson	.449**	1	.195**	.129*	.223**
	Sig(2-tailed)	.000		.003	.048	.001
Neglect from parents	Pearson	.136*	.195**	1	.262**	.246**
	Sig(2-tailed)	.037	.003		.000	.000
School drop outs	Pearson	.200**	.129*	.262**	1	.325**
	Sig(2-tailed)	.002	.048	.000		.000
Poor role models	Pearson	.246**	.223**	.246**	.325**	1
	Sig(2-tailed)	.000	.001	.000	.000	
Political differences	Pearson	-.018	.019	.137*	.133*	.156*
	Sig(2-tailed)	.787	.770	.035	.042	.017
Police siding with criminals	Pearson	.136*	.100	.195**	.121	.276**
	Sig(2-tailed)	.037	.128	.003	.065	.000
Infrastructure	Pearson	.273**	.205**	.096	.036	.272**
	Sig(2-tailed)	.000	.002	.143	.580	.000
Lack of natural surveillance	Pearson	.199**	.153*	.199**	.234**	.320**
	Sig(2-tailed)	.002	.019	.002	.000	.000
Lack of community cohesion	Pearson	.221**	.155*	.238**	.220**	.304**
	Sig(2-tailed)	.001	.018	.000	.001	.000

		Political difference	Police siding	Infrastructure	Lack of natural surveillance	Lack of community cohesion
Drugs	Pearson	-.018	.136*	.273**	.199**	.221**
	Sig(2-tailed)	.787	.037	.000	.002	.001
Unemployment	Pearson	.019	.100	.205**	.153*	.155*
	Sig(2-tailed)	.770	.128	.002	.019	.018
Neglect from parents	Pearson	.137*	.195**	.096	.199**	.238**
	Sig(2-tailed)	.035	.003	.143	.002	.000
School drop outs	Pearson	.133*	.121	.036	.234**	.220**
	Sig(2-tailed)	.042	.065	.580	.000	.001
Poor role models	Pearson	.156*	.276**	.272**	.320**	.304**
	Sig(2-tailed)	.017	.000	.000	.000	.000
Political differences	Pearson	1	.278**	.219**	.236**	.253**
	Sig(2-tailed)		.000	.001	.000	.000
Police siding with criminals	Pearson	.278**	1	.249**	.181**	.215**
	Sig(2-tailed)	.000		.000	.005	.001
Infrastructure	Pearson	.219**	.249**	1	.189**	.282**
	Sig(2-tailed)	.001	.000		.004	.000
Lack of natural surveillance	Pearson	.236**	.181**	.189**	1	.352**
	Sig(2-tailed)	.000	.005	.004		.000
Lack of community cohesion	Pearson	.253**	.215**	.282**	.352**	1
	Sig(2-tailed)	.000	.001	.000	.000	

**correlation is significant at the 0.01 level (2 tailed).
*correlation is significant at the 0.05 level (2 tailed).

The correlation matrix in Table 3 displays different levels of correlation between the different causes of crimes against person.

Generally, the correlations indicate a significant relationship between most of the variables at either 1% or 5% level of significance except the relationships between:

1. Drugs and substance abuse versus racial/political differences
2. Unemployment versus either racial/political difference or police siding with criminals
3. Neglect from parents versus poor infrastructure
4. School drop outs versus either poor infrastructure or police siding with criminal

Table 4. *Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's Test of sphericity.*

KMO value	0.774
Bartlett's Test Approx. Chi-square	330.724
Df	45
Sig.	0.00

The null hypothesis that the correlation matrix is an identity matrix was rejected at 5% level of significance (Bartlett's test of Sphericity has an approximate chi-square that is significant at 45 degrees of freedom), this implies that the correlation in the data set are appropriate for factor analysis. Also, the KMO statistic value of 0.774 reveals that adequate sampling has been used for this analysis [17].

Table 5. *Communalities Initial extraction.*

Cause of crime	Initial	Extraction
Drugs and substance abuse	1.00	0.684
Unemployment	1.00	0.629
Neglect from parents	1.00	0.423
School drop outs	1.00	0.623
Poor role models	1.00	0.455
Political /racial differences	1.00	0.573
Police siding with criminals	1.00	0.420
Poor infrastructure	1.00	0.636
Lack of natural surveillance	1.00	0.391
Lack of community cohesion	1.00	0.427

From table 5, we see that drugs and substance abuse, unemployment, school drop outs and poor infrastructure within the residential estates in Mathare slums were best represented in the common factor space.

Table 6. *Total variance explained by different causes of crimes committed against persons.*

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	2.885	28.854	28.854	2.885	28.854	28.854	1.866	18.662	18.662
2	1.285	12.852	41.706	1.285	12.852	41.706	1.749	17.495	36.157
3	1.089	10.894	52.600	1.089	10.894	52.600	1.644	16.443	52.600
4	.860	8.601	61.201						
5	.777	7.769	68.970						
6	.719	7.193	76.163						
7	.684	6.840	83.004						
8	.636	6.359	89.363						
9	.578	5.775	95.138						
10	.486	4.862	100.000						

Considering the eigenvalue-one criterion and scree plot in in figure 2, it would be reasonable to retain the first three PCs which explain up to 52.6% of the total variability of crimes against persons in Mathare slums.

Figure 2: Scree plot

Table 7. Eigen vectors: Component Factors Estimates.

| Cause of crime	Principal Component		
	1	2	3
Drugs and substance abuse	0.532	-0.616	0.147
Unemployment	0.477	-0.623	0.115
Neglect from parents	0.493	0.036	-0.422
School drop outs	0.490	-0.003	-0.618
Poor role models	0.661	-0.001	-0.134
Political /racial differences	0.419	0.600	0.193
Police siding with criminals	0.503	0.319	0.255
Poor infrastructure	0.527	-0.011	0.598
Lack of natural surveillance	0.591	0.161	-0.124
Lack of community cohesion	0.630	0.172	0.025
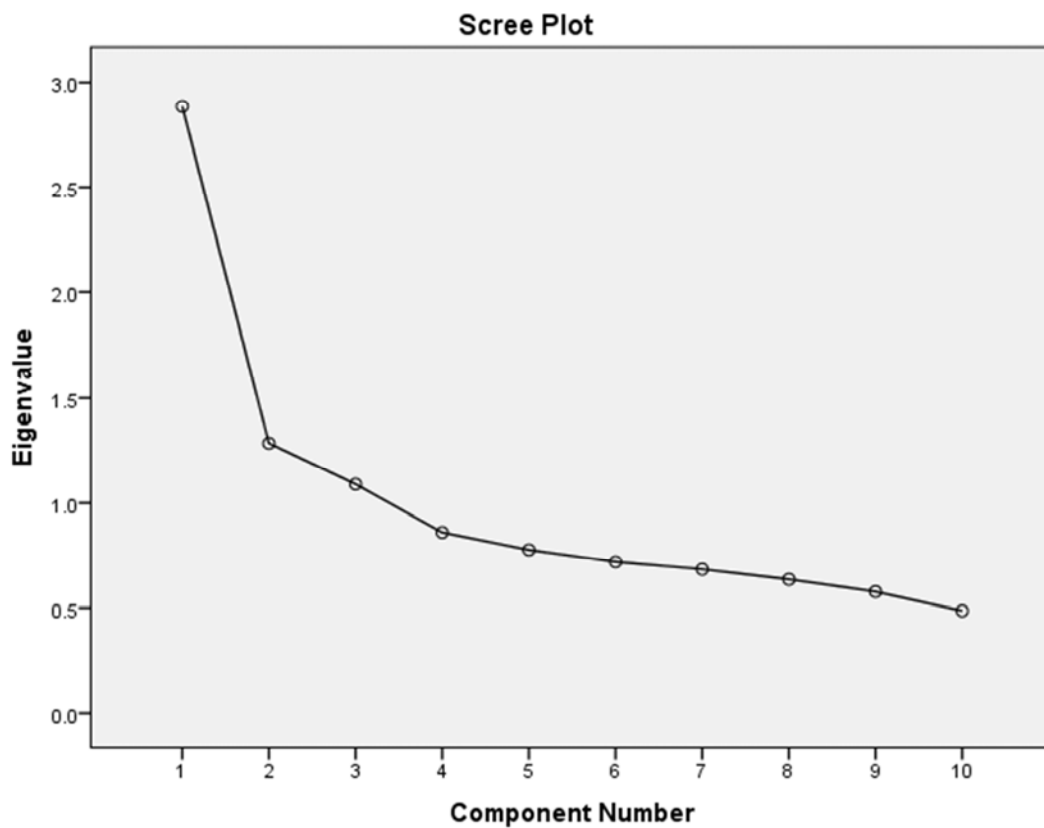
Figure 2. Scree plot.

Table 7 above concentrated on the three PCs that explains 52.6% of the total variability of the data set.

Component 1 has a positive relationship with all the causes of crime but majorly with drugs and substance abuse, poor infrastructure, lack of natural surveillance and lack of community cohesion within the slums.

Component 2 has negative relationship with drugs and substance abuse, unemployment, school drop outs, poor role models and poor infrastructure within the slums otherwise positive relationship with the rest of the factors.

Component 3 has negative relationship with neglect from parents, school drop outs, poor role models and lack of natural surveillance otherwise a positive relationship with the rest of the causes of crimes.

3.2. Causes of Crimes Committed Against Property

Crimes committed against property includes crimes such as robbery, theft/ stealing, pick-pocketing or snatching, house and store breaking.

Table 8. *Descriptive statistics of causes of crimes against property.*

	Cause of crime	Rating (scale: 5-strongly agree, 4-agree, 3-neutral, 2-disagree and 1-strongly disagree)					Number of respondents	mean	Standard deviation
		Strongly agree	agree	neutral	disagree	Strongly disagree			
1.	Drugs and substance abuse	128 (54.5%)	57 (24.3%)	27 (11.5%)	16 (6.8%)	7 (3.0%)	235	4.204	1.079
2.	Unemployment	149 (63.4%)	48 (20.4%)	22 (9.4%)	4 (1.7%)	12 (5.1%)	235	4.353	1.066
3.	Neglect from parents	71 (30.2%)	70 (29.8%)	51 (21.7%)	29 (12.3%)	14 (6.0%)	235	3.660	1.200
4.	School drop outs	99 (42.1%)	63 (26.8%)	50 (21.3%)	13 (5.5%)	10 (4.3%)	235	3.970	1.115
5.	Poor role models	71 (30.2%)	68 (28.9%)	51 (21.7%)	32 (13.6%)	13 (5.5%)	235	3.647	1.201
6.	Political or racial differences	56 (23.8%)	50 (21.3%)	77 (32.8%)	34 (14.5%)	18 (7.7%)	235	3.391	1.212
7.	Police siding with criminals	78 (33.2%)	55 (23.4%)	51 (21.7%)	31 (13.2%)	20 (8.5%)	235	3.596	1.299
8.	poor infrastructure within residential estates	87 (37.0%)	57 (24.3%)	48 (20.4%)	24 (10.2%)	19 (8.1%)	235	3.719	1.280
9.	lack of natural surveillance	70 (29.8%)	73 (31.1%)	58 (24.7%)	22 (9.4%)	12 (5.1%)	235	3.711	1.140
10.	poor community cohesion	78 (33.2%)	60 (25.5%)	60 (25.5%)	22 (9.4%)	15 (6.4%)	235	3.698	1.205

Results in Table 8 reveals that:

1. On average unemployment, drugs and substance abuse and school drop outs are the most important variables that influences crimes against property in Mathare slums. They have the highest mean scores at 4.353, 4.204 and 3.970 respectively. On the other hand, police siding with criminals and political or racial differences have the least influence on crimes against property with mean scores of about 3.596 and 3.396 respectively.

2. The number of respondents who agree and strongly agree that unemployment is a cause of crimes against property is 83.8%. This is followed closely by drugs and substance abuse and school drop outs at 78.8% and 68.9% respectively.

3. The political or racial differences variable has the highest number of undecided respondents of whether it's a cause of crimes against property in Mathare slums. It has 32.8% of undecided respondents. This is followed by existence of poor community cohesion at 25.5%.

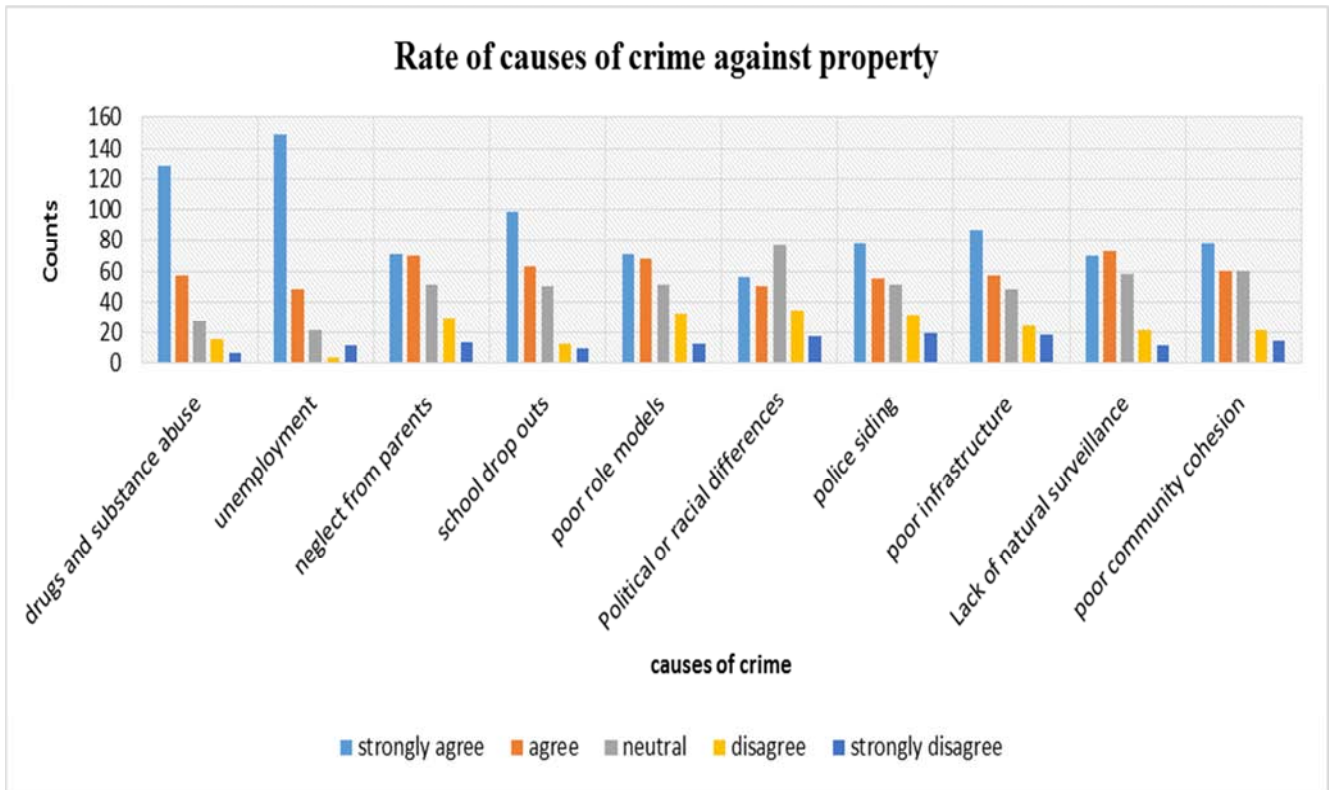A representation of results in table 7 is as shown in figure 3 below.

Figure 3. *Bar graph showing the rate of causes of crimes against property in Mathare slums.*

Table 9. *Correlation matrix for causes of crime against property in Mathare slums.*

		Drugs	Unemployment	Neglect from parents	School drop outs	Poor role models
Drugs	Pearson	1	.394**	.090	.229**	.181**
	Sig(2-tailed)		.000	.168	.000	.005
Unemployment	Pearson	.394**	1	.118	.128*	.064
	Sig(2-tailed)	.000		.071	.051	.325
Neglect from parents	Pearson	.090	.118	1	.325**	.177**
	Sig(2-tailed)	.168	.071		.000	.006
School drop outs	Pearson	.229**	.128	.325**	1	.298**
	Sig(2-tailed)	.000	.051	.000		.000
Poor role models	Pearson	.181**	.064	.177**	.298**	1
	Sig(2-tailed)	.005	.325	.006	.000	
Political differences	Pearson	.174**	.071	.165*	.236**	.236*
	Sig(2-tailed)	.008	.277	.011	.000	.000
Police siding with criminals	Pearson	.203**	-.038	.251**	.213**	.198**
	Sig(2-tailed)	.002	.558	.000	.001	.002
Infrastructure	Pearson	.305**	.245**	.249**	.312**	.208**
	Sig(2-tailed)	.000	.000	.000	.000	.001
Lack of natural surveillance	Pearson	.316**	.215**	.309**	.292**	.212**
	Sig(2-tailed)	.000	.001	.000	.000	.001
Lack of community cohesion	Pearson	.166*	.120	.201**	.181**	.239**
	Sig(2-tailed)	.011	.066	.002	.005	.000

		Political difference	Police siding	Infrastructure	Lack of natural surveillance	Lack of community cohesion
Drugs	Pearson	.174**	.203**	.305**	.316**	.166*
	Sig(2-tailed)	.008	.002	.000	.000	.011
Unemployment	Pearson	.071	-.038	.245**	.215**	.120
	Sig(2-tailed)	.277	.558	.000	.001	.066
Neglect from parents	Pearson	.165*	.251**	.249**	.309**	.201**
	Sig(2-tailed)	.011	.000	.000	.000	.002
School drop outs	Pearson	.236**	.213**	.312**	.292**	.181**
	Sig(2-tailed)	.000	.001	.000	.000	.005
Poor role models	Pearson	.236**	.198**	.208**	.212**	.239**
	Sig(2-tailed)	.000	.002	.001	.001	.000
Political differences	Pearson	1	.182**	.220**	.194**	.116*
	Sig(2-tailed)		.005	.001	.003	.075
Police siding with criminals	Pearson	.182**	1	.258**	.146*	.214**
	Sig(2-tailed)	.005		.000	.025	.001
Infrastructure	Pearson	.220**	.258**	1	.404**	.299**
	Sig(2-tailed)	.001	.000		.000	.000
Lack of natural surveillance	Pearson	.194**	.146*	.404**	1	.344**
	Sig(2-tailed)	.003	.025	.000		.000
Lack of community cohesion	Pearson	.116	.214**	.299**	.344**	1
	Sig(2-tailed)	.075	.001	.000	.000	

Table 9 also presents a fairly high correlation between drugs and substance abuse and unemployment with coefficient of 0.556 which is the highest than the rest of the correlations between other variables

Table 10. *Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's Test of sphericity.*

KMO value	0.784
Bartlett's Test Approx. Chi-square	343.947
Df	45
Sig.	0.00

Table 11. *Communalities Initial extraction*

Cause of crime	Initial	Extraction
Drugs and substance abuse	1.00	0.559
Unemployment	1.00	0.694
Neglect from parents	1.00	0.357
School drop outs	1.00	0.405
Poor role models	1.00	0.338
Political /racial differences	1.00	0.252
Police siding with criminals	1.00	0.395
Poor infrastructure	1.00	0.469
Lack of natural surveillance	1.00	0.462
Lack of community cohesion	1.00	0.291

The KMO measure of 0.784 is again considered adequate

according to Perry et al [17]. Since Bartlett's Test of sphericity has an approximate chi-square that is significant at 45 degrees of freedom with p-value<0.05, we reject the null hypothesis and conclude that the correlation matrix is not an identity matrix and therefore consider the data used suitable

for this type of investigation.

From table 11, we see that drugs and substance abuse and unemployment were best represented in the common factor space.

***Table 12.** Total variance explained by causes of crimes committed against property.*

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	2.989	29.891	29.891	2.989	29.891	29.891	2.384	23.837	23.837
2	1.233	12.327	42.218	1.233	12.327	42.218	1.838	18.381	42.218
3	.948	9.483	51.702						
4	.873	8.728	60.430						
5	.846	8.463	68.893						
6	.771	7.708	76.601						
7	.677	6.775	83.376						
8	.610	6.104	89.480						
9	.593	5.934	95.414						
10	.459	4.586	100.000						

Based on eigenvalue-one criterion and the scree plot in figure 4, we retain the first two PCs which together explain up to 42.2% of the variability of crimes against property in Mathare slums.
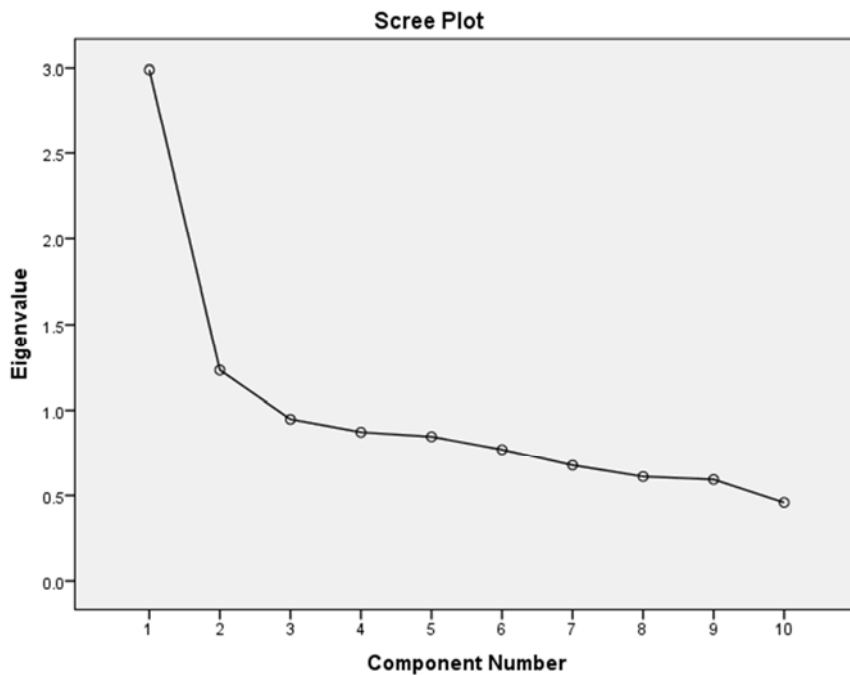
***Figure 4.** Scree plot.*

***Table 13.** Eigen vectors: Component Factors Estimates.*

Cause of crime	Principal Component	
	1	2
Drugs and substance abuse	0.560	0.496
Unemployment	0.393	0.735
Neglect from parents	0.532	-0.272
School drop outs	0.609	-0.185
Poor role models	0.509	-0.281
Political /racial differences	0.454	-0.215
Police siding with criminals	0.468	-0.419
Poor infrastructure	0.676	0.112
Lack of natural surveillance	0.665	0.141
Lack of community cohesion	0.534	-0.076

Table 13 above concentrated on the two PCs that explains 42.2% of the total variability of the data set.

Component 1 has a positive relationship with all the causes of crime but majorly with drugs and substance abuse, school drop outs, poor infrastructure and lack of natural surveillance within the slums.

Component 2 has negative relationship with all the causes of crime except drugs and substance abuse, unemployment, poor infrastructure and lack of natural surveillance within the slums.

4. Conclusion

The main objective of this study was to carry out a statistical analysis of the causes of crimes committed in Mathare slums. The Correlation analysis indicates a fairly strong positive relationship between unemployment and drugs and substance abuse which means that their variables can be used to predict one another. PCA analysis reveals that three PCs (drugs and substance abuse, unemployment and neglect from parents) that explains about 52.6% of the total variability of the causes of crimes against person are suggested to be retained.

Similarly, two PCs (drugs and substance abuse and unemployment) that explains about 42.2% of the total variability of the causes of crimes against property are suggested to be retained.

Generally, the causes of crimes against person and property in Mathare slums are not unique and could be addressed and managed by empowering the residents through programs that can create job opportunities for them as well as arresting the influence of drugs and substance abuse.

Abbreviations

KNBS-Kenya National Bureau of Statistics.
IEBC-independent Electoral and Boundaries Commission.
KMO- Kaiser-Meyer-Olkin measure.
PC-Principal Component.

References

[1] Fajemirokun F., Adewale O., Idowu T., Oyewusi A. and Maiyegun B. (2006). *A GIS Approach to Crime Mapping and Management in Nigeria: A case study of Victoria Island Lagos, CBN*. Journal of Applied Statistics, Vol.3 No.2 49, .www.oicf.org.

[2] Aki Stravra, (September 2002). *Crime on Nairobi: Results of a City wide victim survey*. UN Habitat (safer cities: series 4), Nairobi.

[3] Masese Grace (2007). *Crime and Violence Trends in Nairobi, Kenya*. Case study prepared for Enhancing Urban Safety and Security: Global Report on Human Settlements 2007. UN Habitat 2007.

[4] GOK and UNDP (2010-2013). *Report on overview of crime incidents in Nairobi Region Security Research Information Centre (SRIC)*.

[5] Rev. Fr. Dr. Ndikaru Wa Teresia (2011). *Crime causes and victimization in Nairobi city slums*. International journal of current research, 3 (12) 275-285.

[6] Darkey D. and Kariuki A. (2013). *A study on Quality of Life in Mathare, Nairobi Kenya*. Journal of Human Economic Development. 41 (3) 207-219.

[7] Andvig J. C and Barasa T. (2014). *A political Economy of slum spaces: Mathare valley*. Norwegian Institute of International Affairs. Norway.

[8] Chris Muchwanju, Joel C. Chelule and Joseph Mung'atu (2015). Modelling crime rate using a mixed effect regression model. *American journal of Theoretical and Applied statistics*. 4 (6), 496-503.

[9] Wanjiru M. W and Matsubara K. (2017). Slum toponymy in Nairobi, Kenya. A case study analysis of Kibera, Mathare and Mukuru. *Urban and Regional Planning review*, 4.

[10] Mburu L. W (2014). *Modeling and mapping crime in Eastern Nairobi, Kenya*. GIScience Research Group. University of Heidelberg.

[11] Jolliffe, I. T (2002).*Principal Component Analysis*, 2nd Edition, Springer-Verlag, New York.

[12] Rencher, AC. (2002). Methods of Multivariate Analysis, 2nd edition, John Wiley & Son, New York. Richard, A.J. and Dean, W. W. (2001). Applied Multivariate Statistical Analysis, 3rd edition, Prentice-hall, New Dehli.

[13] Kendall Williams and Ralph Gedeon (2004), *A Multivariate Statistical Analysis of Crime Rate in US Cities.*

[14] Yusuf Bello, Yusuf U. Batsari and Abdullahi S. Charanchi (2014). *Principal Component Analysis of Crime Victimizations in Kastina Senatorial zone*. International journal of science and Technology,3 (4).

[15] Olufolabo O. O.,Akintande O. J., Ekum M. I. (2015). Analysing the Distribution of crimes in Oyo State using Principal Component Analysis. IOSR Journal of Mathematics (IOSR-JM) Vol.11 Issue 3.

[16] Soren, H. (2006). Example of multivariate analysis in R-Principal Component Analysis (PCA).

[17] Perry. R. H., Charlotte, B., Isabell M. and Bob, C. (2004).*SPSS Explained,* ISBN: 0-203-67627-0, Routledge, Taylor &Francis Group, London &New York.