MULTIVARIATE TECHNIQUES IN CRIME DATA ANALYSIS: AN ASSESSMENT OF UTILIZED AND ALTERNATIVE STATISTICAL METHODS

BY

Abdulhameed A. Osi

MSC/SCI/11569/11-12

A THESIS SUBMITTED TO THE SCHOOL OF POSTGRADUATE STUDIES, AHMADU BELLO UNIVERSITY, ZARIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTERS DEGREE IN STATISTICS

DECLARATION

I hereby declare that the work in this thesis titled "**Multivariate Techniques in crime data analysis: An assessment of utilized and alternative Statistical Methods**." was performed by me in the Department of Mathematics under the supervision of Dr. Hussaini Garba Dikko and Professor S. U Gulumbe.

The information derived from the literature has been duly acknowledged in the text and list of references were provided. No part of this work has been presented for another degree or diploma at any institution.

CERTIFICATION

This thesis titled "**Multivariate Techniques in crime data analysis: An assessment of utilized and alternative Statistical Methods**" by Abdulhameed A. OSI meets the regulations governing the award of the degree of Masters of Science in Statistics of the Ahmadu Bello University, and is approved for its contribution to knowledge and literary presentation.

Chairman Supervisory Committee		
Dr. H. G. Dikko	(Signature)	(Date)
Member Supervisory Committee		
Prof. S. U Gulumbe	(Signature)	(Date)
H.O.D Mathematics		
Dr. B. Sani	(Signature)	(Date)
Dean, School of Postgraduate Studies		
Prof. Adebayo A. Joshua	(Signature)	(Date)

DEDICATION

This thesis is dedicated to my beloved parent, Late Alhaji Ado Osi Wudil, Hajiya Binta Abdulkadir and Hajia Aishatu Abubakar who always offer me unconditional love and supports.

ACKNOWLEDGMENTS

All praises be to Almighty Allah for giving me opportunity to complete this work. It is a pleasure to thank the many people who made this thesis possible. Foremost, I would like to express my sincere gratitude to my supervisors Dr. Hussaini Garba Dikko and Professor Shehu U. Gulunbe for their guidance, patience, motivation, enthusiasm and immense knowledge throughout process of this research as well as their interest in its successful completions. I'm highly appreciative for the supports and contributions provided by Mal. Yusif Bello and Mal. Shehu Bala. While they had no responsibilities to me, they willingly listened to my questions and contributed constructive feedback and recommendations. I would like to extend my deepest appreciation to Professor O.E. Asiribo, Dr. Abubakar Yahaya, ACP Gazali A. Azeez (AC CID Katsina Command), CSP Lawal M. Lawal (2I/C CID) and ASP Ahmad U. Kyari (Communication Department NPF Katsina Command) for their suggestion, encouragement and insightful comments. Finally I would like to thank my family: my wife Zainab Muhammad Adam and my Sons Ammar A. Osi and Ahmad A. Osi for their patience, supports and confidence in me throughout my education.

ABSTRACT

The scope of crime and the concern for its prevention/control has grown considerably in the last few years in Nigeria. Therefore, how to discover the variables which have salience in affecting crime rate has become crucial. In this thesis, some of Multivariate Statistical Techniques were utilized on the crime data of Nigeria. Specifically, Principal Component Analysis (PCA) is used to discover the distinct influential variables in the identification of State with high or low crime rate; evaluates and compares the performances of five different Classification Techniques (i.e., Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), K-Nearest Neighbour Analysis (KNN), Classification Trees (CT) and Logistic Discriminant Analysis (LgDA)) in the classification of States of Nigeria as high and low crime rate (unsafe and safe). Each method has unique assumptions about the data, so each may be appropriate for different situations. The results show that four Principal components have been retained using both scree plot and Kaiser's criterion which accounted for 75.024% of the total variation. QDA had the best overall classification performance by classifying 100% of the States correctly, followed by LDA which had only 13.9% apparent error rate. LgDA is recommended to be used when QDA assumption failed while CT are the recommended alternative when LDA's assumptions are not met. Though CT's performance is likely lower than that of LDA, it offers many advantages that make it a useful method, such as its lack of data assumptions.

TABLES OF CONTENTS

Title Page

Declaration……..……..ii

Certification ……..……iii

Dedication……iv

Acknowledgement…….....v

Abstract…….……...vi

Table of contents……vii

List of Figures……..x

List of Tables……..……...xi

List of Appendices……xii

CHAPTER ONE

INTRODUCTION

1.1 Introduction……..…1

1.2 Statement of the problem……..…3

1.3 Aims and Objective of the study…….…….6

1.4 Significance of the study……..…….6

1.5 Scope and limitation…….……8

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction……9

2.2 Concept of crime…….…..…9

2.3 Classification of crime…….....10

2.4 Causes of crimes……..……....10

2.5 Multivariate Analysis in Criminology.……..….12
2.6 Crime analysis using PCA…….……..…15
2.7 Researches on Discriminant Analysis..……..…….……...16

CHAPTER THREE

MATERIAL AND METHOD

3.1 Introduction.…….20
3.2 Description of data…….…..20
3.3 Methodology……..…….22
3.3.1 Principal component Analysis…….…….…..22
3.3.1.1 Derivation of PCs.…….……..….24
3.3.1.2 PCs obtained from the covariance and correlation matrices……..….27
3.3.1.3 Procedure for calculating PCs……..…....28
3.3.2 Quadratic and Linear Discriminant Analysis…..…….……29
3.3.3 The K-nearest neighbor ……..…….……33
3.3.4 Classification tree theory..……..…….35
3.3.5 Logistic discriminant analysis……..……37
3.3.6 Evaluation of Classification Performances……..……39

CHAPTER FOUR

RESULT AND DISCUSSION

4.1 Introduction.……41
4.2 Normality of the data…….…….41
4.3 Principal Component Analysis.……..…….44
4.4 Discriminant Analysis…….……53

CHAPTER FIVE
SUMMARY AND CONCLUSION
5.1 Introduction.……..63

5.2 Conclusion……..…….63
5.3 Recommendation……..……...66
5.4 Improvements for future study……..……..68

References…….……..69

LIST OF FIGURES

Figure	Title	Page
4.1	Normality check samples	43
4.2	The scree plot from correlations	45
4.3	The scree plot from PCs	46
4.5a&b	Score and loading plots of 1st &2nd PCs	48
4.6a&b	Score and loading plots of 1st &3rd PCs	49
4.7a&b	Score and loading plots of 1st &4th PCs	50
4.8a&b	Score and loading plots of 2nd &3rd PCs	51
4.9a&b	Score and loading plots of 2nd &4th PCs	52

LIST OF TABLES

Table	Title	Page
3.1	List of variables…….……..21	
4.1	New variables after transformations……...…….42	
4.2	Correlation matrix for the variables…….…..……..44	
4.3	KMO's and Bartlett's test…….……..……45	
4.4	Eigenvalues of the correlation matrix…….……..…...46	
4.5	Rotated Component Matrix……47	
4.6&4.7	Classification result of LDA on original variables & PCs……..55	
4.8&4.9	Classification result of QDA on original variables & PCs…….……..56	
4.10&4.11	Classification result of KNN on original variables & PCs..…...56 & 57	
4.12&4.13	Classification result of CT on original variables & PCs……..57	
4.14&4.15	Classification result of LgDA on original variables & PCs.……58	
4.16	Summary of performance of all five methods……..59	

LIST OF APPENDICES

Appendix	Title	Page
A1	Dataset before transformation……..	79
A2	Dataset after transformation…….	80
A3	Crime rate of the States……...…….	81

CHAPTER ONE

INTRODUCTION

1.1 Introduction

Crime is one of the persistent problems that bedevil the existence of mankind. Since early days, crime had been a disturbing threat to man, his personality, property and lawful authority. Today, in the modern complex world, the situation is most highly disturbing. Crime started in the primitive day as a simple and less organised phenomenon, and developed today as so complex and organised endeavour. The existence of crime and its attendant problems have therefore spanned the history of mankind.

Multivariate analysis consist of a collection of methods that can be used when several measurement are made on each individual or object in one or more samples. We will refer to the measurements as variables and to the individuals or objects as units (research units, sampling units or experimental units) or observations. There is a wide range of multivariate techniques available but most of them are aimed to simplify the data in some way in order to clarify relationships between variables. The choice of method depends on the nature of the data, the type of problem and the objectives of the analysis. Factor analysis and principal component analysis are methods that can be used to reduce the dimensionality of multivariate data. Cluster analysis seeks to find natural grouping of objects or individuals. Discriminant analysis is a technique designed to clarify the differentiation between groups influenced by the independent variable(s). Other techniques, for example,

multiple regression analysis, aim to explain the variation in two or more independent variables, MANOVA (multiple analysis of variance) an extension of the univariate ANOVA, is used when there are multiple independent variables.

Linear discriminant analysis (LDA) developed by Fisher (1936), is still a commonly used multivariate statistical method for predicting group membership from numeric variables but it makes mathematical assumptions pertaining to data distributions that can invalidate results if violated. It assumes that the data for each group included in the analysis are multivariate normally distributed and that the groups have equal covariance matrices. Other discriminant analysis techniques such as k-nearest neighbor analysis (KNN), quadratic discriminant analysis (QDA), classification trees (CT) and logistic discriminant analysis (LgDA) have since become alternative options. The above mentioned discriminant analyses have assumptions of their own that differ from those of LDA, which may justify their use in situations where LDA's assumptions are violated. For example, QDA assumes multivariate normality, but does not require equal covariances. Therefore, QDA may be a more appropriate method in situations where data have multivariate normal distributions, but unequal covariances. The importance of validating LDA's assumptions has been mentioned by Feldesman (2002). Also states that; many researchers simply proceed with an LDA assuming that the assumptions are not violated.

This study focused on the application of multivariate techniques, specifically principal component analysis (PCA) and discriminant analysis (DA), on some selected factors affection crime management in Nigeria. The works intended mainly to explore the interrelationship between certain crime variables across the States of Nigeria and to

evaluate the assumptions and performance of LDA, QDA, KNN, CT and LgDA in classification of State as safe and unsafe. This very analytical techniques not only can make police department have a sense and the idea of dangerous communities to pay more attention to, in terms of enforcement/control measures, but assist governments in finding out what variables they need to be addressed in order to improve safety and security of the State so as to make it a better place to live as well. The performances, advantages, and disadvantage of the methods will be compared and discussed. Recommendations for future use and research will be made based upon these criteria.\

1.2 Statement of the problem

Crime is often thought of as a moral threat and quite injurious to the society. It afflicts the personality of individual and his property and lessens trust among members of the society (Louis et al., 1981). Crime is indeed an inescapable reality in human life, therefore no national characteristics, no political regime, no system of law, police or justice has rendered a country exempt from crime (Radzinowicz and King, 1977). The causes of crime are multiple and could be traced to bio-genetic factors, such as genetic mutation and heredity (Hooton, 1939), psychological factors, such as personality disorders (Abramson, 1944) and sociological factors, such as learning and environment (Sutherland, 1939). The diverse differences in geographical areas in terms of population density, demographic characteristics, natural vegetations, locations and socio-economic characteristics have rendered crime rate unevenly distributed globally. The disturbing reality is that, it has been observed that the entire world is experiencing high crime rate. The report of the International Crime Victim Survey (ICVS) has confirmed this situation. The report which was conducted on six major world regions including Africa, Asia, Central and Eastern Europe,

Latin America, and Western Europe for the 1989 – 1996 period, has shown that more than half of the urban respondents reported being victimised, at least once, regardless of what part of the world they inhabit (Ackerman and Murray, 2004).

A major problem in the utilization of official crime statistics in Nigeria is that of accessibility, reliability and validity: Official statistics on crime in Nigeria are not easily accessible by the public. Where available, they are not current or up-to-date. For instance, as we said the main avenue for accessing statistics on crime is the *Annual Abstract of Statistics* published by the NBS and the last edition of the publication was the 2011 edition which contained information about year 2010. Up to the mid-1980s the Nigeria Police Force regularly produced and published annual reports that provide meaningful data and statistics. The Report by the police contained data on arrest, cases prosecuted or under investigation, establishment and strength of personnel; discharges and discipline of personnel. At present police do not publish these reports on regular basis.

The starting point here is to ask whether Crime Statistics in Nigeria, as furnished by the police are valid, reliable, and utilizable. *Validity* refers to the extent to which the crime figures are correct representation of actual criminality. *Reliability* refers to the extent to which the crime figures are the result of a consistent measure of the same phenomenon from village to village, town to town and State to State. *Utility* refers to the extent to which the crime figures contribute to the knowledge of crime, improving the understanding, and providing enough information for projection and planning for the future. Crime statistics produced by the Nigeria Police Force are often inaccurate and sometimes incomplete due to administrative and technical problems, includes the statistics sometimes do not include

returns from some state commands that fail to submit their crime returns promptly. And because crime statistics are often now computed on demands by government or the Inspector-General of Police in order to make pronouncements, rather than on an institutionalized basis, the crime statistics produced are often inconsistent as they depend, at any point, on the number of state commands that have submitted returns, the diligence of officials to locate all returns from files, and the competence of compiling officers to ensure accuracy. Therefore, Existing inaccuracies in police crime statistics go beyond the traditional deficiencies, the police record is not valid relative to the actual or "true" volume of crime in the country, this is due to **dark figures** (unreported crimes), **grey figures** (reported but unrecorded crime and manipulation of records to satisfy political and, or institutional interest (as when reported increase or decrease may be advantageous to regime in power or the police force).

At this point one may ask: what factors may be related to safety? Are there some variables we can observe to predict the safety and security of states even without using the actual crime rate statistics? How we construct a baseline lead to classify a particular state? Is there any alternative classification method when LDA's assumptions are violated?

We try to solve the above questions with the help of multivariate statistical techniques based on some variables on States of Nigeria. The potential results are valuable for; (i) police department to increase enforcement and step up patrols in unsafe States (ii) Those living in States with high crime rates to be on the lookout for criminal offenders and (iii) State governments to decrease the crime rate by increasing or decreasing certain variables.

1.3 Aim and Objectives of the Study

The aim of the study is to do pattern recognition by determining which variables greatly affect States' safety and to extract the most important variables, compress and reduce the number of variables by keeping only those important variables, leading to the classification of States according to safeness as well as predicting its safeness, objectives will be:-

1. To determine the number of principal components (PCs) to be retained that would explain sufficient variability of the data set.
2. To evaluate the assumption and performance of five distinct discriminant techniques (ie LDA, QDA, KNN, CT and LgDA) in categorizing States into two groups: high crime rate or low crime rate (safe and unsafe).
3. To provide the recommended alternative techniques when LDA's assumptions are not met.

1.4 Significance of the Study

In Nigeria, research reports of crime surveys and the periodical crime Statistical Abstracts have concentrated mainly in either using descriptive statistics or univariate statistics approaches in their statistical analysis, while little attention is focused on using multivariate statistical approaches. This analysis will help towards better understanding of the dynamic of crime. Considering increasing rise of crime in the country, there is an urgent need to conduct a statistical research that if completed would guide security agencies in

drawing effective action plans and effective deployment of personnel; and in promoting rational decisions by the policy makers. These if achieved, will assist to successively solve many of the complex criminal problems that have bedevilled the country.

Finally, the research work would hopefully provide relevant information that may serve as a guide for those who may wish to conduct further research relating to crime, crime analysis and related issues.

1.5 Scope and Limitation of the Study

The study focused more on the application of principal component analysis, and discriminant techniques. A brief review of basic multivariate analysis was included in the study.

The study covers the thirty seven (37) States of Nigeria including FCT Abuja. The study used secondary data obtained from government agencies such as The Nigeria police Force (NPF), the National Bureau of Statistics (NBS), the National Population Commission (NPC), National Drugs and Law Enforcement Agency (NDLEA) and Independent National Electoral Commission (INEC) based on variable affecting the crime rates in respect of the States of the federation. The reliability of the analysis depends on the accuracy of the data. However, many factors have limited the accuracy of the data. Systematic record keeping of crime data in Nigeria is still very cumbersome for the NPF, although the NPF headquarter is making serious efforts to improve (Ebbe, 1990). The concentration of police stations in the

principal towns and urban areas has rendered the crimes committed in rural areas undetected and or unrecorded.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Upon all the rules being codified into laws by Legislatures, enforced by the police and interpreted by judges, crime continues to exist. What then is that crime? What are the causes of crime? How do we classify crimes? What are the effects/consequences of crime on the well being of a society? What are the statistical tools to analyze the crime? These are some of the questions to be answered in this chapter.

2.2 Concept of Crime

The word 'crime' is from a Latin word *crimen* meaning "accusation" or "fault". There are many definitions of crime. The definition employed by any particular scholar has to do more with his/her area of specialization. Different scholars give different explanations about the occurrence of crime. Who do we blame? Is it the individual who commits the act, the society or both the individual and the society? The question you need to ask is: Do we consider all acts of disobedience as crime? What are the components or elements imagine when we mention the word 'crime'? When can you say crime has indeed occurred or has not?

Legal Definition of Crime

According to Tappan "crime is an intentional act in violation of the criminal law (Statutory and case law), committed without defence or excuse, and penalized by the state

as a felony or misdemeanour" (Tappan, 1960). Crime is a violation of the criminal law, which is subsequently followed by legal punishment. A crime is an act or omission, which attracts sanctions. In studying the offender there can be no presumption that the arrested, arraigned, indicted, or prosecuted persons are criminals unless they are proven guilty beyond all reasonable doubts, of a particular offence. One thing to bear in mind is that, one advantage of a legal definition of crime is that it is narrower and less ambiguous than a social definition of crime. It is a well held view that only a behaviour that violates the criminal law by definition can be regarded as a crime.

2.3 Classification of Crime

The classification of crime differs from one country to another. In Nigeria, the police classification of crimes also depends on what law prescribed. In Nigeria Police Abstract of Statistics (NPACS), offences are categorised in to four main categories:

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

2.4 Causes of Crimes

Criminal behaviour cannot be explained by a single factor, because human behaviour is unformed by complex interaction between genetic, environmental, social, psychological and

cultural factors. Different types of crimes are being committed by different types of people, at different times, in different places, and under different circumstances (Danbazau, 2007). Here we discuss some of the causes of crime.

1. Biogenetic factors: Criminologists are of the opinion that criminal activity is due to the effect of biologically caused or inherited factors. According to Lombroso (1911), a criminal is born, not made; that criminals were the products of a genetic constitution unlike that found in the non-criminal population.
2. Social and Environmental factor (Sutherland, 1939): The environment is said to play significant role in determining criminal behaviour. Factors within the environment that mostly influence criminal behaviour includes poverty, unemployment, corruption, urbanisation, family, moral decadence, poor education, technology, child abuse, drug trafficking and abuse, architectural or environmental design.

Oyebanji (1982) and Akpan (2002) have attributed the current crime problem in Nigeria to urbanisation, industrialisation and lack of education. Kutigi (2008) has said that the factors of crime in Nigeria are poverty and ignorance which is at the same time is the opinion of many Nigerians (Azabuike, 2007). In another dimension, according to Ayoola (2008), lack of integrity, transparency and accountability in the management of public funds, especially at all levels of government have been identified as the factors responsible for the endemic corruption that has eaten deep into the fabric of the Nigerian society over the years.

2.5 Multivariate Analysis in Criminology

Different schools and researchers in criminology study different features relating to the concept, nature, prevention and control of crime. For instance, the location in space and time is an essential feature of any criminal event. The two components are so closely linked

that they are considered simultaneously. This has led to much focus upon the spatial domain called the “hot spot” of crime (Sharman, *et al* 1989; koper, 1995; Mazerolle, 1995).

Techniques such as Multivariate Analysis of Variance (MANOVA), Discriminant Analysis (DA), Factor Analysis (FA), Principal Component Analysis (PCA), Cluster Analysis (CA) {McGurk, *et al* 1986), Log-Linear Analysis (Crowley and Adrain 1992), Multi-Dimensional Scaling (Smith, *et al* 1986), and the more sophisticated Structural Equation Analysis (Liu and Kaplan 1999) were routinely used by criminal justice scholars in the examination of crime phenomenon.

The development of Geographical Information System (GIS) and the popularity of crime mapping among criminal justice practitioners have further emphasized the attention on the spatial dimensions of crime (Buckley, 1996; Harries, 1999).

Several research studies on longitudinal impact of demographic shifts on crime trends have been carried over the years. Criminologists such as Bachman (1992), willliams and Flewelling (1988) and Bailey (1984) using Multivariate Analysis of Variance (MAVOVA) involving 3 or more population means, have hypothesized that differences in homicide rate, and by extension to crime rate, were caused by social conditions such as level of poverty, high unemployment rate or social disorganization that exist in those area. These hypotheses suggest that crime is related to material deprivation and weak social relationship.

Williams and Gedeon (2004), in an application of multivariate statistical analysis on crime in U.S cities, analyzed 14 variables on 100 U.S cities using Principal Component

Analysis, Factor Analysis and Discriminant Analysis to categorize the cities into two groups: high crime rate and low crime rate.

Recent studies by the Crime Mapping Research Center at the National Institute of Justice categorize hot spot detection and analysis methods. These techniques have been classified as follows (Harries, 1999): visual interpretation, choropleth mapping, grid cell analysis, spatial autocorrelation, and cluster analysis.

Andreeseu in his research multivariate analysis is been used on cross-sectional data from the European social survey Round-3 2006/2007 conducted on representative samples of persons aged 15 and over in Romania and Hungary, this comparative study explore the relationships between victimization, trust in the police and fear of crimes.

Hung and Nguyen applied principal component analysis to 25 standard offence groups for 600 cities. The purpose is to derive a small number of 'component' or 'factors' which are highly correlated and representative of the original larger group of offences from this they were able to derive new groupings of cities through their crime pattern which would not have been evident previously.

Greebgerg and Kessler estimate multivariate panel models for the effect of clearance rates and a vector of cocioeconomice control variables on index crimes, using a sample of 98 US cities for the years 1964-1970.

Scalora and Garbin studied the impact of treatment involvement and subsequent recidivism; variables studied include demographic, criminal history, offence related and treatment progress. Recidivism was assessed through arrest data. Multivariate analysis suggests that recidivism is significantly related to quality of treatment involvement, offender demographic, offence characteristics, and criminal history. Successfully treated offenders were significantly less likely to subsequently reoffend.

Fang (2011) in his research PCA is used to discover the distinct influential variables in the identification of a community with high or low crime rate and construct a baseline to classify a community as safe and unsafe using discriminant analysis.

Karofi (2005) carried out a multivariate analysis on drug abuse and criminal behaviour in Penang, Malaysia, to establish that between crime variable and drug abuse, peer group, poverty, broken homes, urbanization and migration variables, poverty is the best predictor at inmates' involvement in criminal behaviour.

In contrast to the deterrence theory which might lead to prediction that a longer period in detention might lead to a decrease rate of future offending, the review of the empirical literature by Hood and Sparks (1990) on variance estimation using two population means, concluded that length incarceration had no clear effect on the post release behaviour. Kohfeld and Sprague (1990) found by using Multiple Regression analysis that the faster the police respond to crime the lower the crime rate may be in a given community.

2.6 Crime Analysis using PCA

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. If some group of measures constitutes the scores of numerous variables, the researcher may wish to combine the score of the numerous variables in to smaller number of super variables to form the group of the measures. This problem mostly happens in determining the relationship between some socio-economic factors and crime incidences. PCA uses the correlations among the variables to develop a small set of components that empirically summarised the correlations among the variables.

Elizabeth (2004) in her effort to determine a model that provides a general relationship between violence and some structural measures. In one of the measures, the concentrated disadvantage index, the PCA of the structural measures supported the combination of the following measures: percentage of residents who are African American, percentage of residents living below the poverty line, percentage of families receiving public assistance, percentage of residents who are unemployed, and percentage of families headed by a single parent with children under 18. Similarly, in a study to examine the statistical relationship between crime and socio-economic status in Ottawa and Saskatoon, the PCA was employed to replace a set of variables with a smaller number of components, which are made up of inter-correlated variables representing as much of the original data set as possible (Exp, 2008).

Principal component analysis can also be used to determine the overall criminality. When the first eigenvector shows approximately equal loadings on all variables then the first

PC measures the overall crime rate. In PrintCom (2003) for 1977 US crime data, the overall crime rate was determined from the first PC, and the same result was achieved by Hardle and Zdenek (2007) for the 1985 US crime data.

2.7 Researches on Discriminant analysis

Giles and Elliot (1962) were the first anthropologists to use multivariate statistical analysis in an attempt to estimate race; their inquiry focused on the identification of American Blacks, American Whites, and American Indians. They believed the cranium provided the best indication of race. Thus, they conducted an LDA on 225 males and 225 females using eight cranial variables: glabello-occipital length; maximum cranial width; basion-bregma height; maximum bi-zygomatic diameter; prosthion-nasion height; basion-nasion; basion-prosthion; and nasal breadth. The analysis resulted in the generation of four discriminant functions, two for each sex, which predicted American White versus American Black and American White versus Native American. Giles and Elliot (1962) reported high classification rates (re-substitution – 86.7% males, 89.8% females; holdout – 79.8% males, 86.6% females), but subsequent tests of their functions failed to match the reported performance (Birkby, 1966; Snow *et al.*, 1979; Ayers *et al.*, 1990; Fisher and Gill, 1990) The functions' performance may not have been ideal, but the method of analysis was pioneering (Iscan, 1988), it set the stage for future studies of ancestry. Multivariate statistics, especially LDA, are still relied on for assessing human craniometric variation and ancestry estimation.

Almost a decade after Giles and Elliot (1962) W. W. Howells published the results of an extensive investigation of global human craniometric variation (Howells, 1973). This study sought to establish a basis for comparison of skulls by constructing a sample of total cranial variation and subjecting it to multivariate analyses. Howells's main question was whether

differences in cranial form among populations are based upon the same factors as differences within populations. His goals were to find variables that could differentiate populations in terms of cranial shape and to describe how shape differs across populations. His sample contained 1,652 crania of individuals from 17 populations distributed across five major geographic regions: Europe; Africa; Asia; the Pacific; and America. Howells collected data comprising 70 measurements and angles and used them in a variety of multivariate analyses, such as LDA and cluster analysis. Though Howells (1973) acknowledged that his sample missed Southeast Asia and large populations, such as China and India, his results still showed a geographic patterning to human cranial variation. Howells published another influential study in 1989, where he explored whether sets of characteristics can distinguish a given population from others and whether there are traits specific to geographic regions. This time he specified six world regions, which he identified as: Europe; Africa (sub-Saharan); the Far East (Japan and China); Australo-Melanesia; Polynesia; and America. He sampled 18 populations, each region represented by three groups. If distinguishing characteristics were apparent, Howells wanted to know whether they reflected ancient separations of regional groups or recent microevolution. He conducted his analysis by subjecting the C-scores (describing cranial shape, while controlling for size) of 57 cranial measurements for each individual to cluster analysis and Q-mode analysis. His results revealed patterned regional variation in cranial shape, supporting his 1973 findings. However, Howells (1989) noted that the observed differences between populations were not great. He also noted that no individual population or regional group was articularly distant or distinguished from the others. Aside from his results concerning patterned geographic variation, Howells (1973;1989) also contributed useful methodology and a great deal of data that is still used in contemporary discriminant analyses evaluating craniometric variation and population affinity (Sauer and Wankmiller, 2009).

Discriminant analysis describes a category of statistical methods that assess the degree to which it is possible to separate two or more groups of observations given a set of variables (Rencher, 2002; Manly, 2005; Huberty and Olejnik, 2006). LDA is a parametric method that assumes that the data for each group have a multivariate normal distribution. It also assumes homogeneity of variance, that the covariance matrices for all groups in the reference sample are equal (Rencher, 2002; Manly, 2005; Huberty and Olejnik, 2006). While these two phenomena can be difficult to test, a researcher or practitioner using LDA must evaluate the assumptions (Ousley and Jantz, 2012).

Merce r(2013) in "Nonparametric Discriminant Analysis in Forensic Ancestry Estimation: An Assessment of Utilitzed and Alternative Statistical Methods. " evaluates and compares the performances of four discriminant analysis techniques (i.e., linear discriminant analysis, quadratic discriminant analysis, k-nearest neighbor analysis, and classification trees) in forensic ancestry estimation using craniometric variables.

Discriminant analysis describes a category of statistical methods that assess the degree to which it is possible to separate two or more groups of observations given a set of variables (Rencher, 2002; Manly, 2005; Huberty and Olejnik, 2006). LDA is a parametric method that assumes that the data for each group have a multivariate normal distribution. It also assumes homogeneity of variance, that the covariance matrices for all groups in the reference sample are equal (Rencher, 2002; Manly, 2005; Huberty and Olejnik, 2006). While these two phenomena can be difficult to test, a researcher or practitioner using LDA must evaluate the assumptions (Ousley and Jantz, 2012). QDA method requires multivariate

normal distribution, but does not assume homogeneity of variances, KNN is non-parametric method that does not assume multivariate normality, CT are discriminant method that make no distributional assumption or variance assumptions, and LgDA is a partially parametric method falling between parametric and nonparametric, in which instead of making assumption about the distribution of the data within each group, LgDA is based on the assumption that the likelihood ratio of the group have exponential form.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

The causes and origins of crime are numerous and complicated and they compared and differ a lot from area to area. When talking about factors which may have impact on the crime rate, everyone can think of several variables. Fortunately, variables affecting crime have been the subjects of investigation to many disciplines historically. We can therefore summarize those known factors impacting on crime volume and crime type, occurring from place to place, as follows:

1. Population density and level of urbanization

2. Composition of the population, particularly percentage of youth

3. Modes of transportation system

4. Economic condition

5. Cultural and Educational factors

6. Climate factors

7. Effective strength of law enforcement agencies

8. Citizens' attitudes toward crime

3.2 Description of data

The data contains 13 variables that fall under several general characteristics categories such as population characteristics, economic characteristics, social characteristics, 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 States in 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), 2011 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 37 entries; each entry represents the information of a particular state in Nigeria. Table 3.1 are the list of all the variables and what they stand for.

Table 3.1: List of Variables

POP	POPULATION
DENS	POPULATION DENSITY
SEX	SEX RATIO
AVRG	AVERAGE HOUSEHOLD SIZE
ILLT	ILLITRACY RATE (YOUTH)
MOD	MODERATE HOUSEHOLD INCOMES
PVR	POVERTY RATE
UNR	UNEMPLOYMENT RATE
VOT	VOTERS TURNOUT
POL	POLICE STRENGHT
GDP	GROWTH DOMESTICS PRODUCT
ARRT	DRUG OFFENDERS ARRESTED
SAI	DRUG SEIZURE/ARREST INDEX

3.3 Methodology

Five techniques were evaluated for their performance predicting group membership from craniometric variables: linear discriminant analysis; quadratic discriminant analysis; k-nearest neighbour analysis; classification trees; and logistic discriminant analysis. These five

methods fall under the general category of discriminant analysis. Manly (2005:105) states, "The problem that is addressed with discriminant analysis is the extent to which it is possible to separate two or more groups of individuals, given measurements for these individuals on several variables." Depending on the research goals, discriminant analysis can be performed to explore differences among groups of observations or to establish rules for predicting group membership (Rencher, 2002; Huberty and Olejnik, 2006). This research is of the latter goal and is primarily concerned with classification accuracy. The linear, quadratic, k-nearest neighbour and Logistic analyses were performed in Stata11, while the classification tree analysis was performed in SPSS.

3.3.1 Principal Component Analysis

Principal component analysis (PCA) is probably the most popular multivariate statistical technique, and it is used by almost all scientific disciplines. It is also likely to be the oldest multivariate technique. In fact, its origin can be traced back to Pearson (1901) or even Cauchy (1829, see Grattan-Guinness, 1997, p. 416), or Jordan (1874, and also Cayley, Silverster, and Hamilton, see Stewart, 1993; Boyer and Merzbach, 1989, for more) but its modern instantiation was formalized by Hotelling (1933) who also coined the term *principal component*.

Having a large number of variables in a study makes it difficult to decipher patterns of association. Variables sometimes tend to repeat themselves. Repetition is a sign of multicolinearity of variables, meaning that the variables may be presenting some of the same information. Principal Components Analysis simplifies multivariate data in that it reduces the dimensionality of the data. It does so by using mainly the primary variables to

explain the majority of the information provided by the data set. Analysis of a smaller number of variables always makes for a simpler process.

Simply stated, in principal components analysis we take linear combinations of all of the original variables so that we may reduce the number of variables from p to m, where the number m of principal components is less than p. Further, the method allows us to take the principal components and use them to gain information about the entire data set via the correlation between the principal components and the original variables. Matrices of correlations or loadings matrices show which principal component each variable is most highly associated with. The first principal component is determined by the linear combination that has the highest variance. Variance measures the diffusion of the data. After the first principal component is obtained, we must determine whether or not it provides a sufficient amount of or all of the information displayed by the data set. If it does not provide adequate information, then the linear combination that displays the highest variance accounted for after the first principal component's variation is removed is designated as the second principal component. This process goes on until an ample amount of information/variance is accounted for. Each principal component accounts for a dimension and the process continues only on the remaining dimensions. Designating a dimension as a principal component often reveals information about correlations between remaining variables which at first was not readily available.

3.3.1.1. Derivation of Principal Components

Let *X* be a vector of *p* random variables, and that the variances of the *p* random variables and the structure of the covariance or correlations between the *p* variables are of

interest. Unless *p* is small, or the structure is very simple, it will often not be very helpful to simply look at the *p* variances and all the $\frac{1}{2}p(p-1)$ correlations or covariance. An alternative approach is to look for a few (< *p*) derived variables that preserved most of the information given by these variances and correlations or covariance. This explained the basic idea of PCA. This is achieved by transforming to a new set of variables, the PCs, which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables (Jolliffe, 2002).

Algebraically, PCs $Y_1, Y_2, ..., Y_p$ are particular linear combinations of the *p* random variables $X_1, X_2, ..., X_p$. Geometrically, these linear combinations represent a new coordinate system obtained by rotating the axes of the original system (the *X*'s). The new axes represent the direction of maximum variability.

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

Consider the linear combinations

$$\begin{aligned} Y_1 &= \alpha_1' X = \alpha_{11} X_1 + \alpha_{12} X_2 + ... + \alpha_{1p} X_p \\ Y_2 &= \alpha_2' X = \alpha_{21} X_1 + \alpha_{22} X_2 + ... + \alpha_{2p} X_p \\ &\vdots \\ &\vdots \\ Y_p &= \alpha_p' X = \alpha_{p1} X_1 + \alpha_{p2} X_2 + ... + \alpha_{pp} X_p \end{aligned} \qquad (3.20)$$

Where $\alpha'_i = (\alpha_{i1}, \alpha_{i2}, ..., \alpha_{ip})$ for i=1, 2,……..,p

of the element of ***X***, where α_j is a vector of *p* components $\alpha_{j1}, \alpha_{j2}, ..., \alpha_{jp}$.

Then

$$Var(Y_j) = \alpha_j' \Sigma \alpha_j \qquad j = 1,2,...,p \qquad (3.21)$$

$$Cov(Y_j, Y_k) = \alpha_j' \Sigma \alpha_k \qquad j,k = 1,2,...,p \qquad (3.22)$$

The PCs are those *uncorrelated* linear combinations $Y_1, Y_2, ..., Y_p$ whose variances in (3.11) are as large as possible (Richard and Dean, 2001).

Proceeding in this way and writing in matrix notation, the result for a random variable *X* with $E(X) = \mu$ and $Var(X) = \Sigma = \Gamma \Lambda \Gamma'$ is the PC transformation which is defined as

$$Y = \Gamma' X \quad \text{or} \quad Y = \Gamma'(X - \mu) \qquad (3.23)$$

In finding the PCs we concentrate on the variances. The first PC is the linear combination with maximum variance. That is, it maximizes $Var(Y_1) = \alpha_1' \Sigma \alpha_1$. It follows that α_1 cannot be taken simply for maximization, since $Var(Y_1) = \alpha_1' \Sigma \alpha_1$ can be increase by multiplying any α_1 by some constant. We need to put some conditions of choosing α_1 to be a coefficient vector of unit length, so that $\alpha_1' \alpha_1 = 1$.

We therefore define

First PC = Linear combination $\alpha_1' X$ that maximizes $Var(\alpha_1' X)$ subject to

$$\alpha_1' \alpha_1 = 1.$$

Second PC = Linear combination $\alpha_2' X$ that maximizes $Var(\alpha_2' X)$ subject to

$$\alpha_2' \alpha_2 = 1 \text{ and } Cov(\alpha_1' X, \alpha_2' X) = 0.$$

and, so on, so that at the *k* th stage

k th PC = Linear combination $\alpha_k' X$ that maximizes $Var(\alpha_k' X)$ subject to

$$\alpha_k' \alpha_k = 1 \text{ and } Cov(\alpha_j' X, \alpha_k' X) = 0 .$$

$$j = \alpha_1' X, \alpha_2' X, ..., \alpha_{k-1}' X$$

Up to *p* PCs could be found, but we have to stop after the *q*th stage (*q* ≤ *p*) when most of the variation in ***X*** have been accounted for by *q* PCs.

Let **Σ** be the covariance matrix associated with the random vector $X' = \lfloor X_1, X_2, ..., X_p \rfloor$. Let **Σ** have the eigenvalue – eigenvector pairs $(\lambda_1, \alpha_1), (\lambda_2, \alpha_2), ..., (\lambda_p, \alpha_p)$ where $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_p \geq 0$. The *j*th PC is given by

$$Y_j = \alpha_j' X = \alpha_{j1}' X_1 + \alpha_{j2}' X_2 + ... + \alpha_{jp}' X_p, \qquad j = 1,2,..., p$$

then

- the variance of a PC is equal to the eigenvalue corresponding to that PC,

$$Var(Y_j) = \alpha_j' \Sigma \alpha_j = \lambda_j \qquad j = 1,2,..., p$$

- the covariance between two different PCs is zero

$$Cov(Y_j, Y_k) = \alpha_j' \Sigma \alpha_k = 0 \qquad j \neq k$$

- The total variance in a data set is equal to the total variance of PCs.

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

3.3.1.2. Principal Components obtained from the Covariance and Correlation Matrices

Consider the covariance matrix of a bivariate data $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$, where $\sigma_{12} = \sigma_{21}$ and $\sigma_{22} \supset \sigma_{11}$, assume σ_{11} =1 and σ_{22} = 100 (a very large difference), and the derived correlation matrix

$$\rho = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}, \qquad \text{where } r_{12} = r_{21}$$

Because of its large variance, X_2 will completely dominates the first PC determine from **Σ**. Moreover, this first PC explains a larger proportion of the total population variance as

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \psi_{X_1}.$$

When the variables X_1 and X_2 are standardized, however, the resulting variables contributes equally to the PCs determine from ρ, so that $\rho_{Z_1,Y_1} = \rho_{Z_2,Y_1}$ and the corresponding proportion of the first PC of the total population variance explained is

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = \psi_{Z_1}$$

Thus, $\psi_{X_1} \neq \psi_{Z_1}$, the proportion of variance accounted for by the components of ρ differ from the proportion for **Σ**.

Since the entries of the covariance and correlation matrices are different, then the coefficients (eigenvectors) of the PC obtained from ρ differ from those obtained from **Σ**,

and therefore $\alpha_j' X \neq \alpha_j' Z$; the PCs derived from ρ are different from those obtained from **Σ** (Rencher, 2002).

3.3.1.3. Procedure for Calculating PCs

For a random vector $X' = \lfloor X_1, X_2, .., X_p \rfloor$, the corresponding standardized variables are $Z' = \lfloor Z_1, Z_2, ..., Z_p \rfloor$ so that *Cov*(Z) = ρ (the correlation matrix of *X*).

We denote the matrix of correlation between *p* variables by

$$\rho = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \qquad (3.25)$$

and the vector of the coefficients (weights or loadings) on the *p* variables for the j th component by

$$\alpha_j = \begin{bmatrix} \alpha_{j1} \\ \alpha_{j2} \\ \vdots \\ \alpha_{jp} \end{bmatrix} \qquad j = 1,2,..., p\,.$$

The problem of determining the vectors α_j which maximize (1) the variance accounted for by the first component, (2) the variance accounted for by the second component, orthogonal to the first, etc., The solution for α_j can be solve by this equation

$$\left(\rho - \lambda_j I\right)\alpha_j = 0$$

in which ***I*** is the (*p*×*p*) identity matrix, λ_j's are the characteristics roots or eigenvalues of ρ and the α_j's are the associated eigenvectors.

The solution of such an equation for λ_j and α_j may be obtained through the following steps:

i. Obtained the characteristics equation of **Σ**, that is $\left|\rho - \lambda_j I\right| = 0$ which leads $\rho - \lambda_j I$ to a polynomial equation in λ_j. Solve this equation for the λ_j's.

ii. For each eigenvalue λ_j, obtained in step i., write out the matrix $\rho - \lambda_j I$.

iii. Compute the $adj\left(\rho - \lambda_j I\right)$

iv. Any column of $adj\left(\rho - \lambda_j I\right)$ is an eigenvector associated with λ_j. (Linderman, 1980).

Having obtained the eigenvectors, we simply determine our PCs as $Y_j = \alpha_j' Z$ for $j = 1,2,...,p.$

3.3.2 Quadratic and Linear Discriminant Analysis Theory

Discriminant Analysis is a method used in statistics and pattern recognition to find a combination of features which characterize or separate two or more classes of objects or events. The resulting combination may be used as a classifier to assign objects to previously defined classes. What discriminant analysis does is to assign objects to previously defined groups. The process of classification, which is what we utilize, defines guidelines such that when followed one can determine which group an object belongs to. In our analysis, we are

only concerned with the case of two groups, X_{g1} and X_{g2}, where $X_{g1} \equiv MN(\mu_1, \sum_1) and\ X_{g2} \equiv MN(\mu_2, \sum_2)$ are two multivariate normal populations.

We distinguish between our two groups based upon the values of our random variables $X^T = [X_1, X_2, ..., X_P]$, where each group's values for each variable differ to some degree. Each group has a population consisting of the values of its variables defined by a probability density function $f_1(x) or\ f_1(x)$. The above mentioned guidelines are developed via a training sample. Two regions are formed, R_1 and R_2. The training sample splits the majority of the original sample into two known or correctly classified (by characteristics) regions and then each region R_1 and R_2 is associated with the group, π_1 and π_2 respectively. The remaining sample, n minus the size of training sample, is called the test sample. This is used to test the validity of the classification rule formed by the training sample.

Measurements of all objects of one class k are characterized by a probability density function $f_k(X)$ which is seldom known. And there might be some prior knowledge about the probability of observing a member of class k, the prior probability π_k with $\pi_1 + \pi_2 + ... + \pi_k = 1.$ to estimate $f_k(X) and\ \pi_k,$ a training sample is used. Most often applied classification rules are based on the multivariate normal distribution.

$$f_k(X) = f(X \setminus k) = \frac{1}{(2\pi)^{p/2} \left|\sum_k\right|^{1/2}} e^{-\frac{1}{2}(X-\mu_k)^T \Sigma_k^{-1} (X-\mu_k)} \qquad (3.31)$$

Where $\mu_k \; and \; \Sigma_k$ are the class k population mean vector and covariance matrix. Under such assumption, the probability that one object with given vector $X = (X_1, X_2, ..., X_p)$ to belong to class k can be calculated by the formula below

$$p(k \Big/ X) = \frac{f(X/k)\pi(k)}{\sum_{k=1}^{K} \pi(k) f(X \Big/ k)} \alpha \, f(X \Big/ k)\pi(k) \tag{3.32}$$

Taking the logarithm of Equation 3.32 will lead to the discriminant function

$$d_k(X) = (X - \mu_k)^T \Sigma_k^{-1} (X - \mu_k) + \log|\Sigma_k| - 2\log \pi_k \tag{3.33}$$

and the classification rule

$$d_{\hat{k}}(X) = \min d_k(X) \Leftrightarrow \max \; p(k \mid X) \tag{3.34}$$

The rule described in Equation 3.33 and Equation 3.34 is called Quadratic Discriminant Analysis (QDA). When a special case that all *k* class covariance matrices are identical $\Sigma_k = \Sigma$, the discriminant function can be simplified to Equation 3.35 which is called the Linear Discriminant Analysis (LDA).

$$d_k(X) = 2\mu_k^T \Sigma^{-1} X - \mu_k^T \Sigma^{-1} \mu_k - 2\log \pi(k) \tag{3.35}$$

Therefore, our first step in performing Discriminant Analysis is to check to see whether or not our covariance matrices, Σ_1 and Σ_2, from our two group model are equal. We

check the equality of our covariance matrices in order to know if we could apply linear Discriminant Analysis or Quadratic Discriminant Analysis. We check the equality of the covariance matrices by testing the null hypothesis H_0: $\Sigma_1 = \Sigma_2$ against H_1: $\Sigma_1 \neq \Sigma_2$. To test the null hypothesis, we evaluate the pooled unbiased estimate of the common covariance matrix under H_0, which is given by

$$S_p = \frac{1}{n_1 + n_2 - 2}\left\{\sum_{i=1}^{2}(n_i - 1)S_i\right\}, \qquad (3.2.0)$$

Where N_i is the sample size of group i and S_i is the i^{th} sample covariance matrix. After evaluating S_p, we calculate the test statistic for the equality of the covariance matrices, which has a chi-square (χ^2) distribution and is equal to M/c.

$$M = \left\{\sum_{i=1}^{2}(n_i - 1)\right\}\ln(\det(S_p)) - \left\{\sum_{i=1}^{2}(n_i - 1)\ln(\det(S_i))\right\}$$

$$1/c = 1 - \frac{2p^2 + 3p - 1}{6(p+1)}\left[\left(\sum_{i=1}^{2}\frac{1}{n_i - 1}\right) - \frac{1}{n_1 + n_2 - 2}\right]$$

Incorrect classification sometimes does occur in discriminant analysis due to the fact that the characteristics or variables of the two populations may not always be readily distinguishable. Some contributing factors to misclassification are incomplete knowledge of future performance, exhaustion of the object required for faultless information, and the event of information not being readily accessible. The guidelines followed for classification should minimize the frequency of a misclassification occurring. When determining guidelines

one must look at factors such as prior probabilities and the cost of misclassification. To minimize the expected cost of misclassification (ECM) one would want the following to hold for each region:

$$R_1 : \frac{f_1(x)}{f_2(x)} \geq \left(\frac{c(1/2)}{c(2/1)} \right) \left(\frac{p_2}{p_1} \right)$$
$$R_2 : \frac{f_1(x)}{f_2(x)} < \left(\frac{c(1/2)}{c(2/1)} \right) \left(\frac{p_2}{p_1} \right),$$

Where c is the cost that an object is misclassified and $p_1 \, and \, p_2$ are the prior probabilities for $\pi_1 \, and \, \pi_2$. The left side of the inequalities is known as the density ratio. Under a multivariate normal population, the rule for assigning an object to either group becomes:

Allocate $\mathbf{x}_0$ to X_{g1} if

$$(\mu_1 - \mu_2)^T \Sigma^{-1} x_0 - \frac{1}{2}(\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 + \mu_2) \geq \ln \left[\left(\frac{c(1/2}{c(2/1)} \right) \left(\frac{p_1}{p_2} \right) \right].$$

Allocate $\mathbf{x}_0$ to X_{g2} if otherwise. Another method used to attain optimal classification would be to minimize the total probability of misclassification (TPM):

$$TPM = \alpha = 2\Phi \left(-\frac{1}{2} \Delta p \right),,$$

Where Φ(z) is the cumulative distribution function of the standard normal and

$$\Delta p = \sqrt{(\underline{\mu}_1 - \underline{\mu}_2)^T \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2).}$$

3.3.3 The K-Nearest Neighbour (KNN)

KNN was the first non-parametric discriminant analysis developed (Rencher, 2002). It does not assume multivariate normality, but it does assume homogeneity of variance. KNN differs from the above methods in that it does not find functions to discriminate groups, but classifies an observation based upon the group membership of a number, *k*, of its nearest neighbors. The distance between two observations is calculated as Khattree and Naik, 2000):

$$d^2(x_1, x_2) = (x_1 - x_2)\mathrm{V}^{-1}(x_1 - x_2) \qquad 3.0$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are observation vectors and **V** is the covariance matrix. KNN posterior probabilities of group membership are calculated by the formula (Khattree and Naik, 2000):

$$p_i = \frac{\pi_i(k_i/n_i)}{\sum_{i=1}^{g}\pi_i(k_i/n_i)} \qquad 3.2$$

where π_i is the prior probability of belonging to group *i*, k_i is the number of observations amongst the k nearest neighbors belonging to group *i*, and n_i is the total number of observations in group *i*. In the case where *k*=1 an observation would be classified into the group that its single closest neighboring observation belongs to. When *k*>1 an observation in question is classified into the group that is most common among its *k* nearest neighbors (Rencher, 2002; Huberty and Olejnik, 2006). Thus, KNN analysis assumes that observations from the same group are close to one another in multivariate space.

The researcher chooses the value of *k* and if *k* is too large then observations that are more distant from the one being classified, and theoretically less similar, will be considered. Researchers should seek to keep *k* small so that very close observations are used for the classification (Huberty and Olejnik, 2006).

A decision must be made as to the value of k. optimal choice of k for KNN is not an exact science. With two groups, k should be chosen as an odd integer to avoid ties. Rencher (2002, 319) cites the research of Loftsgaarden and Quesenberry (1965), which suggests that an optimal k is $\sqrt{n_i}$, where n_i is a typical group size. Rencher also suggests running with several different values of k and choosing the one that gives the best error rate. McLachlan (2004) cites Enas and Choi (1986), which suggests that when there are two groups of comparable size that k should be chosen approximately between $N^{3/8}$ or $N^{2/8}$, where N is the number of observations. In this analysis *k*=3 was used.

3.3.4 Classification Trees

Classification trees are non-parametric procedures that classify observations by repeatedly partitioning the data into subsets through a series of decisions (Breiman *et al.*, 1984). The goal is to create final subsets that are homogenous with respect to the group or class variable. Classification trees start with all of the observations grouped together in one node, referred to as the root node. All of the predictor variables are evaluated to determine which can be used to split the root node into groups that best separate the classes. The best split is determined by evaluating a measure of impurity that quantifies the class makeup of

each node. The impurity measure is maximized when all of the classes are equally mixed in a node and minimized when a node contains only one class (Breiman *et al.*, 1984). The chosen split is the one that most greatly reduces the impurity.

To choose a split rule the tree algorithm evaluates each variable one-by-one. It determines a value on each variable that would provide the best split. Next, it compares the splits for each variable to determine which does best and selects that variable (Breiman *et al.*, 1984).

A classification tree can construct binary or multinomial splits. This analysis uses binary splits. Therefore, a left and a right node are created each time a node is split. The new nodes' impurities are evaluated as well as another set of split decisions for each node. If the node's impurity can be reduced, then the split process is repeated. However, if the impurity cannot be significantly improved, then the splitting is stopped and the node becomes a terminal node. The terminal node is assigned a class equal to that of the class with the highest proportion of observations in the node. All of the observations in a terminal node are predicted as belonging to the class associated with that node. A new observation is classified by subjecting it to the decisions, or split rules, of the tree until the observation reaches a terminal node.

Breiman *et al.* (1984) discuss a number of advantages to using classification trees. Classification trees do not require multivariate normality and can be applied to any data structure. This makes them more generally applicable and removes the need to evaluate the distribution of the data for each group in the analysis. The procedure does not require homogeneity of variance. Its conditional decisions made upon individual nodes do not

require equal or similar data distributions. Classification trees inherently perform variable selection in their evaluation and selection of optimal splitting decisions. They are very robust to outliers and are even robust to misclassifications of observations in the training data. They can perform analyses in the presence of missing data and have a number of built-in procedures for handling missing values (Breiman *et al.*, 1984; Ding and Simonoff, 2010). Finally, the procedure is easy to use and provides output that can be easily interpreted and understood.

On the other hand, classification trees have a number of difficulties and drawbacks. First, the researcher must pay attention to the growth of the tree and make sure that it is not overly complex, or over-fit. Breiman *et al.* (1984) recommend growing the tree to its maximum and then selectively pruning it, recombining some of the later nodes. Though some programs, such as SAS Enterprise Miner, have functions built in that evaluate tree size and performance based on cross-validation or a holdout sample, this issue still requires attention when constructing a decision tree. Second, there are a variety of rules that can be selected for evaluating splitting. There are various measures of impurity and other methods, such as misclassification rates, that can be used to determine splits and the researcher must decide which is most appropriate. Third, for groups that can be separated by a linear function or some other combination of variables, it may take a very complex decision tree with many splits in order to achieve the same performance. In this case it is preferable to use another method, such as LDA. Lastly, divergent sample sizes can affect decision tree performance. Potential decisions are evaluated based upon how much they reduce the impurity, average performance across all ten runs is reported. Breiman *et al.* (1984) state that this method performs well with classification trees.

3.3.5 Logistic discriminant analysis

Albert and Lesaffre (1986) explain that logistic discriminant analysis is a partially parametric method falling between parametric discrimination methods such as LDA and QDA and nonparametric discrimination methods such as k-nearest-neighbour (KNN) discrimination and classification trees, provide a good explanation of logistic discriminant analysis. Instead of making assumptions about the distribution of the data within each group, logistic discriminant analysis is based on the assumption that the likelihood ratios of the groups have an exponential form. Multinomial logistic regression provides the basis for logistic discriminant analysis and can handle binary and continuous regressors, hence logistic discriminant analysis is also appropriate for binary and continuous discriminating variables.

Let g be the number of groups, n_i the number of observations for group i, and q_i the prior probability for group i. Let x denote an observation measured on p discriminating variables. For consistency with the discriminant analysis literature, x will be a column vector, though it corresponds to a row in your dataset. Let $f_i(x)$ represent the density function for group i, and let $P(x/G_i)$ denote the probability of observing x conditional on belonging to group i. Denote the posterior probability of group i given observation x as $P(G_i/x)$. With Bayes' theorem, we have

$$P(G_i \Big/ x) = \frac{q_i f_i(x)}{\sum_{j=1}^{g} q_j f_j(x)}$$

Substituting $P(x/G_i)$ for $f_i(x)$, we have

$$P(G_i \Big/ x) = \frac{q_i P(x/G_i)}{\sum_{j=1}^{g} q_j P(x/G_i)}$$

Dividing both the numerator and denominator by $P(x/G_g)$, we can express this as

$$P(G_i / x) = \frac{q_i L_{ig}(x)}{\sum_{j=1}^{g} q_j L_{jg}(x)}$$

where $L_{ig}(x) = P(x/G_i) / P(x/G_g)$ is the likelihood ratio of x for groups i and g.

This formulation of the posterior probability allows easy insertion of the multinomial logistic model into the discriminant analysis framework. The multinomial logistic model expresses $L_{ig}(x)$ in a simple exponential form

$$L_{ig}(x) = \exp(a_{oi} + a_i' x)$$

And that an individual is classified in group i if $a_{oi} + a_i' x > 0$ the parameters can be estimated by maximum likelihood using an interactive procedure.

3.3.6. Evaluation of classification performances

Several parameters can be used for the quality estimation of classification models, both for fitting and validation purposes (Frank and Todeschini, 1994). Of course, these parameters are related to the presence of errors in the results (objects assigned to the wrong classes), even if errors can be considered with different weight on the basis of the classification aims. All the classification indices can be derived from the confusion matrix, which is a square matrix with dimension GXG, where G is the number of classes. The diagonal elements of confusion matrix represents the correctly classified objects, while the off diagonal elements represent the objects erroneously classified.

By looking at the confusion matrix (built on fitting or validated outcomes), we can have an idea on how a classification model is performing; of course, some more informative

indices can be derived in order to synthesize this information. First, the non-error rate (NER) can be defined as follows:-

$$NER = \frac{\sum_{g=1}^{G} n_{gg}}{n}$$

Where n is the total number of objects. The NER (also called accuracy or classification rate) is the simplest measure of the quality of classification model and represents the percentage of correctly assigned objects. The NER complementary index is called the error rate (ER); it is the percentage of wrongly assigned objects and is defined as;

$$ER = \frac{n - \sum_{g=1}^{G} n_{gg}}{n} = 1 - NER$$

NER and ER can simply describe the performance of a model.

There are also indices related to the classification quality of single class. The sensitivity (Sn_g) describes the model ability to correctly recognize objects belonging to the g^{th} class and is defined as;

$$Sn_g = \frac{n_{gg}}{n_g}$$

If all objects belonging to the g^{th} class are correctly assigned ($n_{gg} - n_g$), Sn_g is equal to 1. The specificity (Sp_g) characterizes the ability of the g^{th} class to reject the objects of all other classes and is defined as;

$$Sp_g = \frac{\sum_{k=1}^{G} (n_k^{'} - n_{gk})}{n - n_g} \qquad for\ k \neq g$$

CHAFTER FOUR

RESULT AND DISCUSSION

4.1 Introduction

In this chapter we run Principal Component Analysis on all 13 variables and five Discriminant Analyses for the normalized and non-normalized; original data and PCs. It should be noted that PCA is sensitive to the scaling of the variables. In our case, variables are measured on different scales, so we need to use the correlation matrix instead of covariance matrix to have the unit variance of data.

4.2 Normality of the data

Often, before doing any statistical modelling, it is crucial to verify whether the data at hand satisfy the underlying distribution assumptions. Multivariate normal distribution is one of the most frequently made distributional assumptions when using multivariate statistical techniques, e.g. Quadratic and Linear Discriminant Analyses. Also, from an important property of multivariate normal distribution, we know that if $X = (X_1, X_2, ..., X_P)$ follow the multivariate normal distribution, then its individual components $X_1, X_2, ..., X_P$ are all normally distributed. Therefore, we need to examine normality of each Xi to guarantee that $X = (X_1, X_2, ..., X_P)$ is multivariate normally distributed.

Here, Quantile-Quantile plot (QQ plot) is used to assess normality of data though there are more formal mathematical assessment methods. The reason is that with a large

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

Drawing QQ plot of each variable reveals that 6 out of the 13 variables approximately follow normal distribution, they are: MOD, PVR, UNR, VOT, SEX and AVRG. We try different forms of transformation on the remaining 7 variables to obtain the substitute variables which perform better on normality. Table 4.1 list the specific forms used to get the new variables. Figure 4.1 is the QQ plot of POP, DENS and GDP in terms of both before-transformation and after-transformation. We can see the effectiveness of data transformation method because data tends to be normally distributed after transformation.

Table 4.1: New variables which are after transformation

New Variables
LOG(POP)
LOG(DENS)
LN(ILLITRACY)
MODERATE
POVERTY

UNEMP

SEXRATIO

AVERAGE

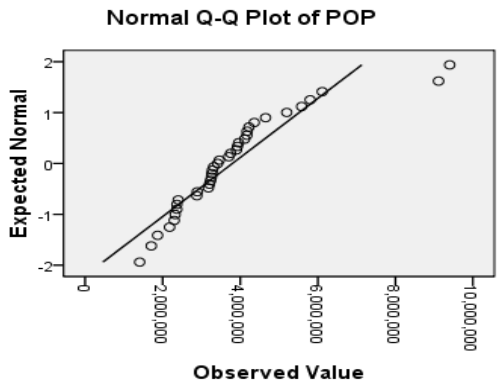
VOTERS

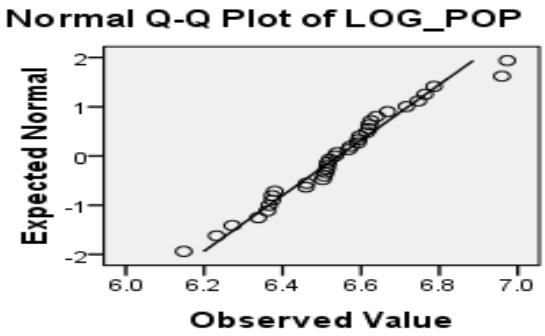
SQRT_LOG(POLICE)

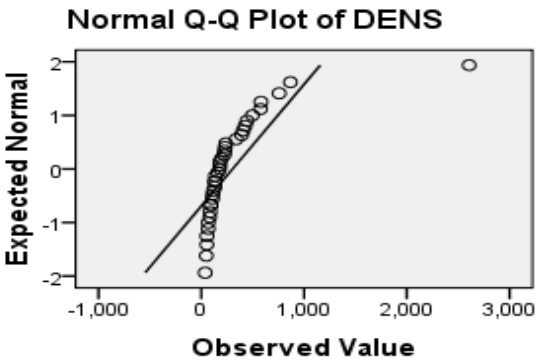
LN_LN_LOG(GDP)

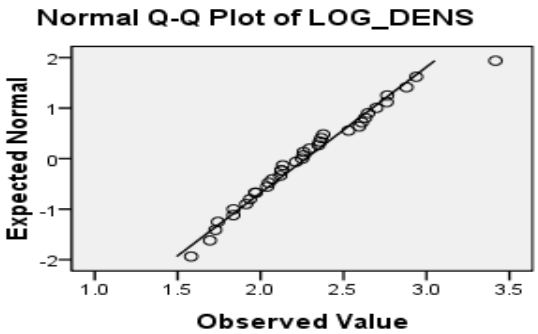
LOG(ARREST)

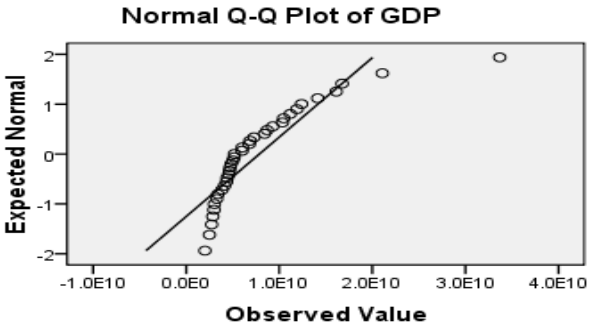
LOG(SAI)











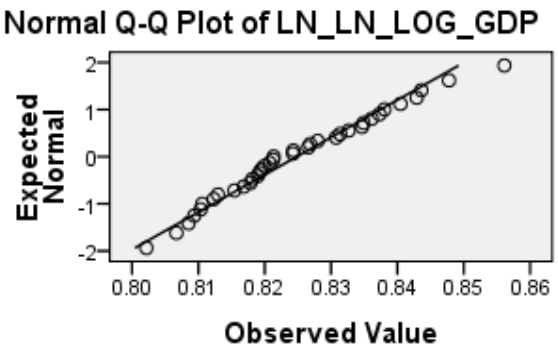


Figure 4.1: Normality check samples

4.3 Principal component analysis

Table 4.2: Correlation Matrix for the variables

	POP	DEN	SEX	AVRG	ILLT	MOD	PVR	UNR	VOT	POL	GDP	ARRT	SAI
POP	1.00	0.407	-0.003	0.046	-0.005	0.107	-0.017	-0.016	-0.030	-0.459	0.665	0.524	0.096
DENS	0.407	1.000	-0.014	-0.483	-0.755	-0.080	-0.366	-0.414	0.158	-0.155	0.603	0.462	0.039
SEX	-0.003	-0.014	1.000	0.258	0.000	0.008	-0.339	-0.065	0.276	0.276	0.118	0.202	-0.124
AVRG	0.046	-0.483	0.258	1.000	0.643	-0.163	0.533	0.522	0.249	-0.149	-0.388	0.042	-0.437
ILLT	-0.005	-0.755	0.000	0.643	1.000	0.265	0.574	0.473	-0.205	-0.118	-0.500	-0.163	-0.131
MOD	0.107	-0.08	0.008	-0.163	0.265	1.000	-0.100	-0.089	-0.623	0.209	0.007	0.037	0.300

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

Determinant = 9.84E-005

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

Table 4.3: KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.527
Bartlett's Test of Sphericity	Approx. Chi-Square	284.479
	df	78
	Sig.	.000

Table 4.3 shows the result of KMO measure of sampling adequacy and Bertlett's test of sphericity. 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 0.527 which is greater than 0.5: so, we should be confident the PCA is appropriate for these data.

Bartlett's measure tests the null hypothesis that the original correlation matrix is an identity matrix. For PCA to work we need some relationships between variables and if the R-matrix were an identity matrix then all correlation coefficients would be zero. Therefore we want this test to be significant. For these data, Bartlett's test is highly significant ($p < 0.001$), and therefore PCA is appropriate.

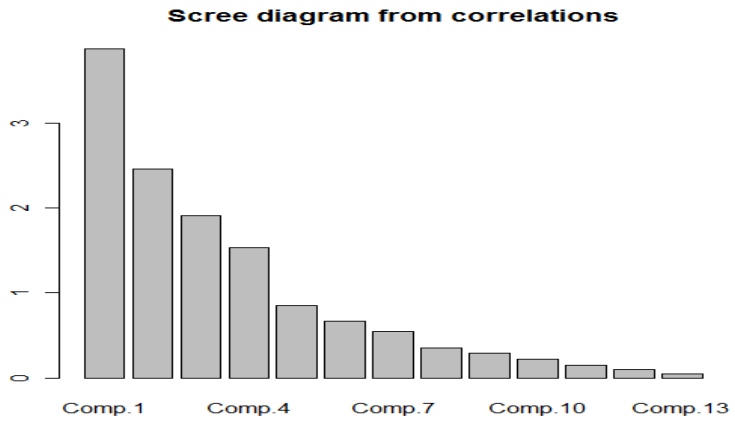


Figure 4.2: The scree Plot from correlations

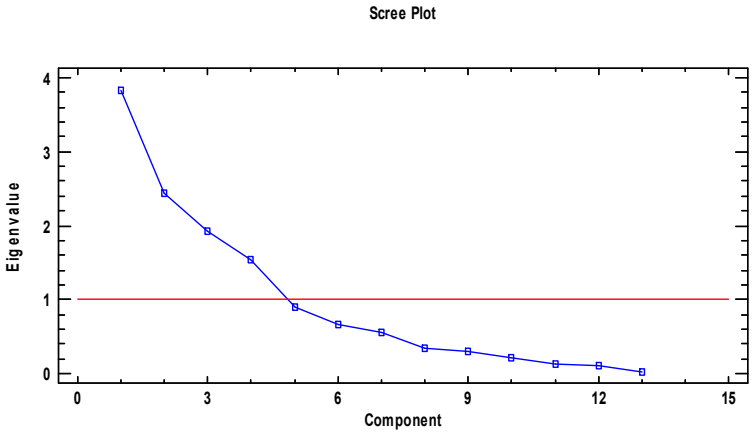


Figure 4.3: The Scree Plot of the PCs

Table 4.4: Eigenvalues of the Correlation Matrix

Component		*Percent of*	*Cumulative*
Number	*Eigenvalue*	*Variance*	*Percentage*

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

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 four PCs that explain up to 75.024 percent of the total variability of the data set as shown in the Table 4.4. We can also use the scree plot to choose the number of PCs to retain, as shown in Figure 4.2 and 4.3 the "elbow" appears at the 5th PC, therefore the first four PCs should be retained which account for 75.024 % of the total variance, as a valuable reduction in dimensionality.

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

Table 4.5 Rotated component matrix

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

The next step is to look at the content of variables that load onto the same factor to try identifying common themes. In Table 4.5, 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, VOTERS has relative large absolute value, hence PC2 denotes as political affiliation factor. Similarly, view PC3 as population factor, PC4 as factor related to gender and police

strength. However, it is kind of difficult to define PC1 in which ILLTRACY, AVERAGE, DENS, UEMNR, POVERTY and GDP all have relative large absolute value of greater than 0.3 level, which indicates that interpretation of PCs is one major disadvantage of PCA.

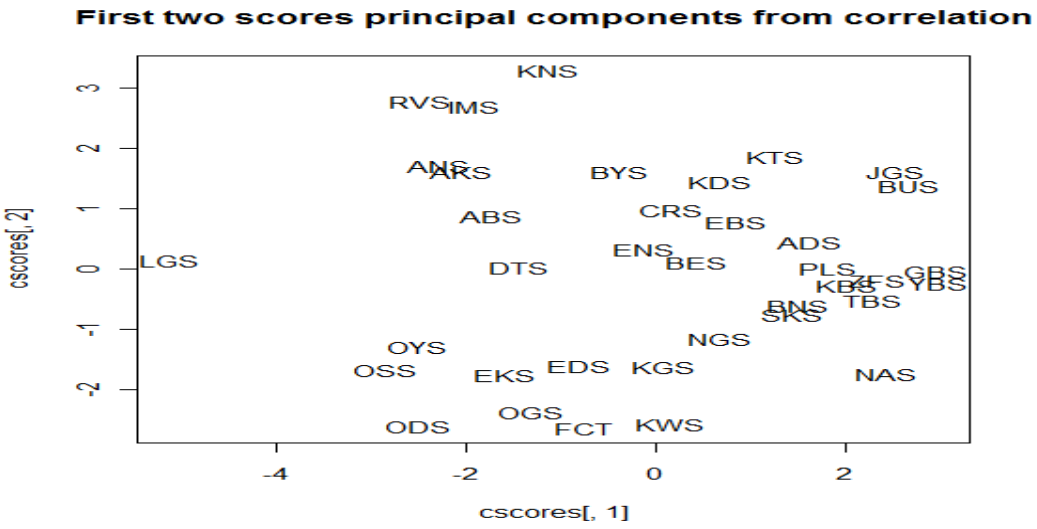


Figure 4.5a score plot of first and second PCs

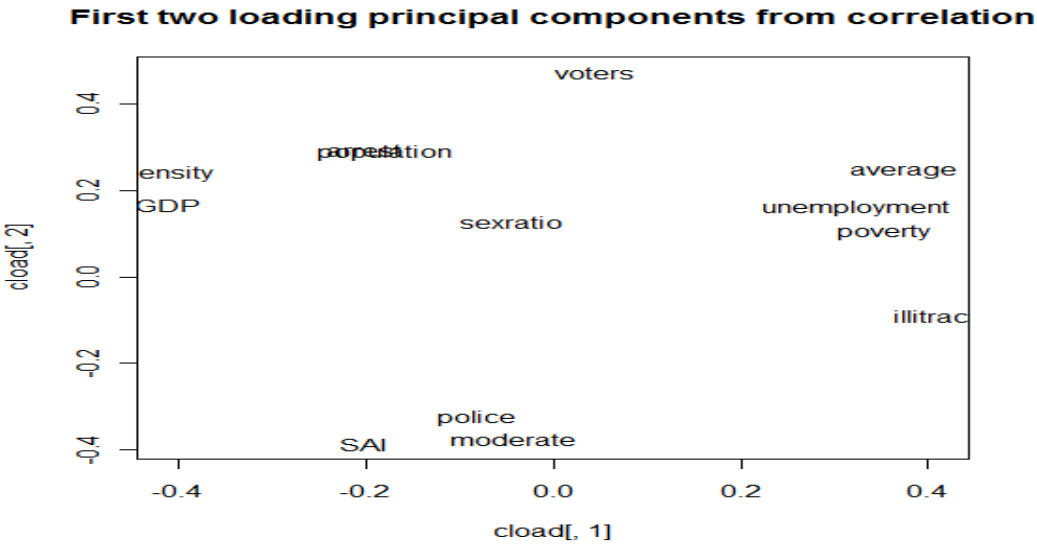


Figure 4.5b Loading plot of first and second PCs

From figure 4.5a and figure 4.5b we can see from loading and scores plot the State with combine prevalence of Population Density and GDP is Lagos State (LGS).

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

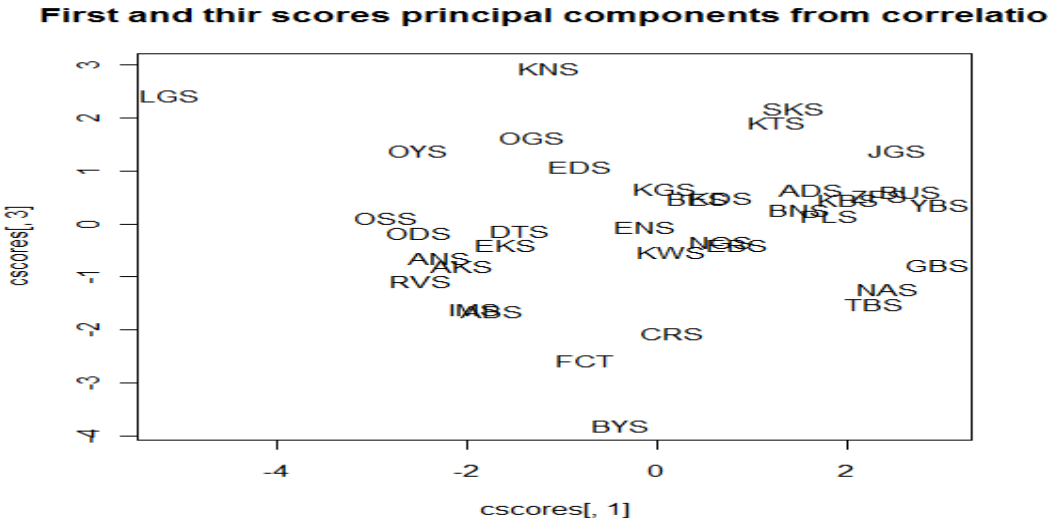


Figure 4.6a Scores plot of first and third PCs

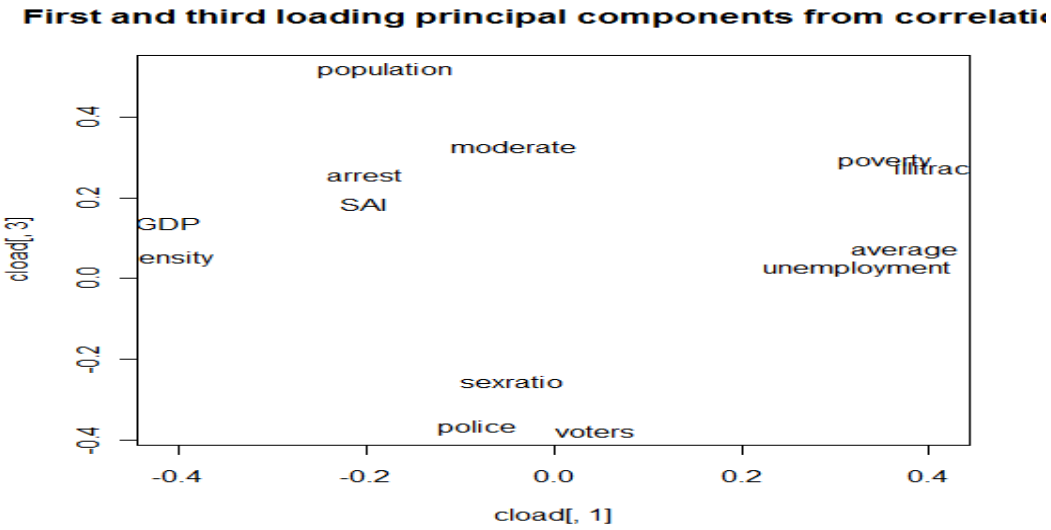


Figure 4.6b Loading plot of first and third PCs

Figure 4.6a and figure 4.6b and shows that Kano (KNS) and Lagos (LGS) States are the most populous States, while Sokoto (SKS), Katsina (KTS), Gombe (GBS), Jigawa (JGS), Bauchi (BUS) and Adamawa (ADS) States have the predominance of Poverty Rates

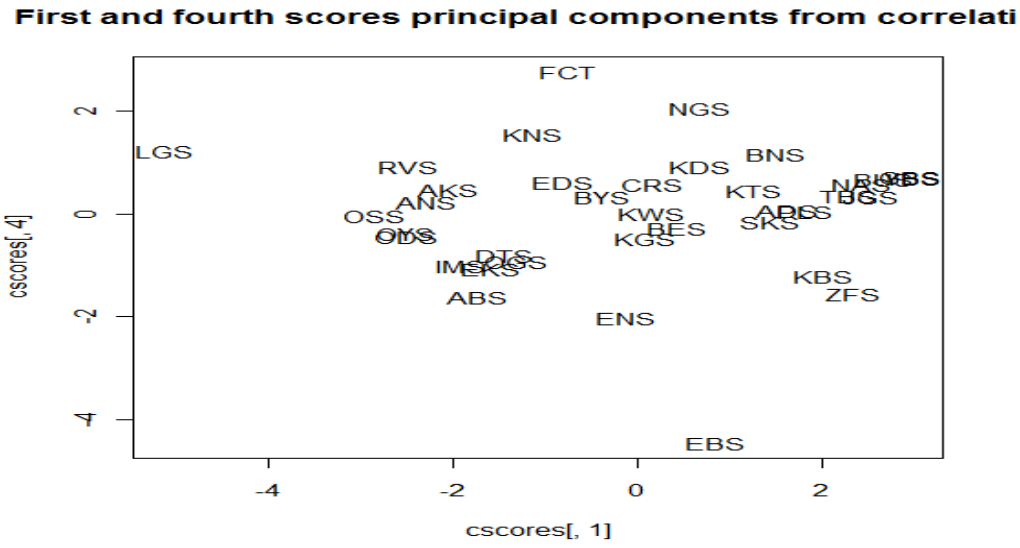


Figure 4.7a Scores plot of first and fourth PCs

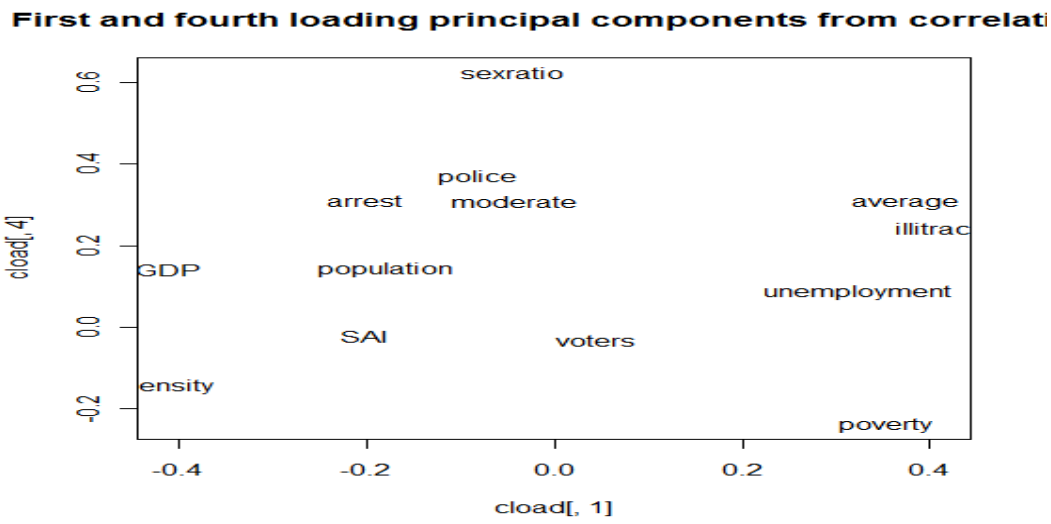


Figure 4.7b Loading plot of first and fourth PCs

Figure 4.7a and figure 4.7b show that FCT Abuja has the higher male to female Sex ratio

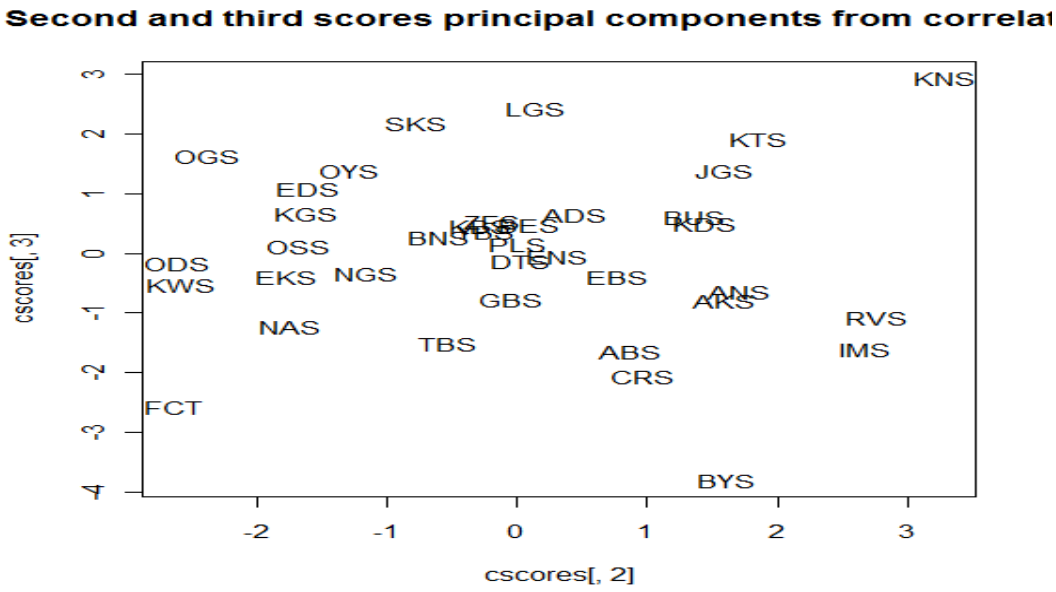


Figure 4.8a Scores plot of second and third PCs

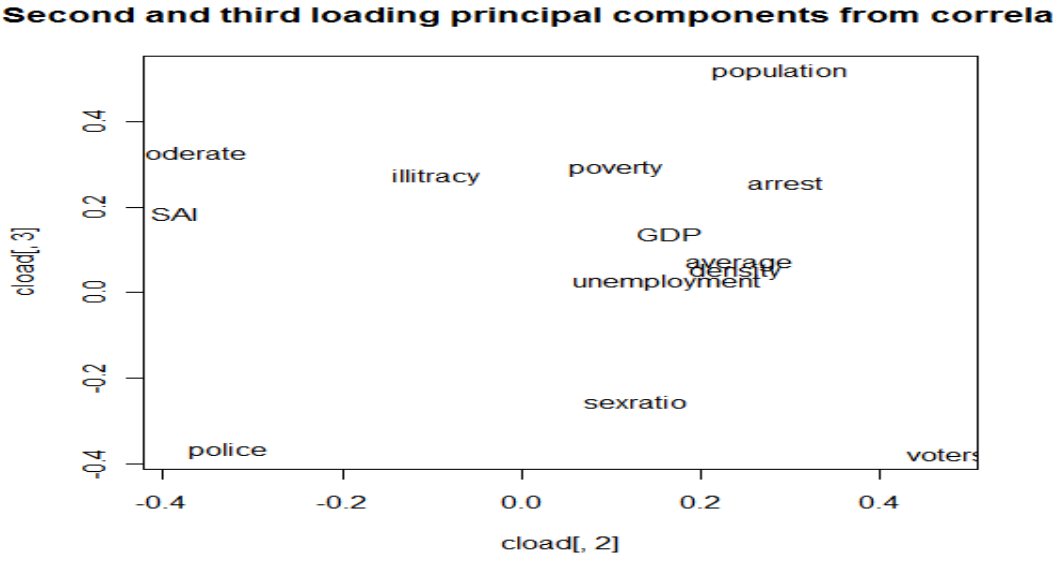


Figure 4.8b Loading plot of second and third PCs

Figure 4.8a and 4.8b shows Kano (KNS), Katsina (KTS) and Jigawa (JGS) States have a tendency towards number of drug offender arrested, while Bayelsa (BYS), Abia(ABS), Cross River (CRS), Imo (IMS) and Rivers States (RVS) with higher turnout for 2011 presidential

election, at the same time Ondo (ODS) State has greater quantity of drug Seizure Arrest Index.

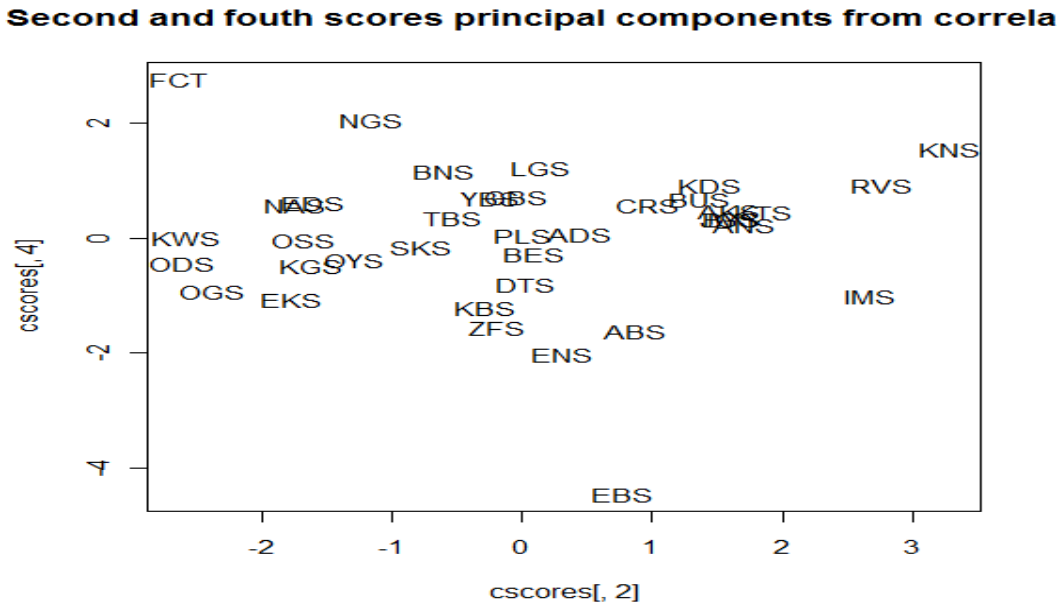


Figure 4.9a Scores Plot of second and fourth PCs

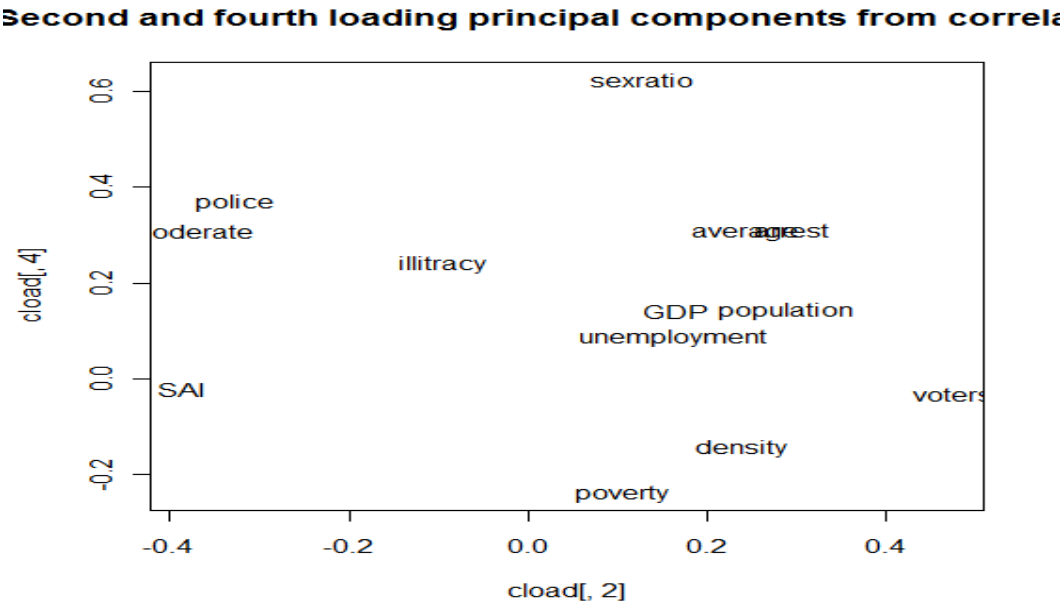


Figure 4.9a Loading Plot of second and fourth PCs

From figure 4.9a and figure 4.9b FCT Abuja and Nasarawa (NSR) State have high proportion of number of police officers per 100,000 populations.

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

4.4 Discriminant Analysis

The original data is a data set that has 13 variables for 37 states of Nigeria. Here, one more response variable named CRIME will be introduced. This variable is a record of number of crimes per 100,000 people for two years. We can see Table A.3 of Appendix A for reference. So our objective is to describe the 37 states as safe or not by using Discriminant Analysis on the known information. One thing may be noticed that the period is not consistent with other 13 variables, but we just assume it is acceptable, because our focus is on the DA method.

Afterwards, the States have been classified as either 'high' or 'low' in crime rate based on the State's average, which is 117.24 crimes per 100,000 people. Assign "1" to states which are above the average and "0" to those below the average. Thus, we have 19 "0s" and 18 "1s" for the 37 states.

The data contains only 37 entries, which is felt too small data size to get the accurate classification rule if it is split into training sample and testing sample, the whole 37 by 13 data matrix is used to form the baseline of classification and test its effectiveness on these 37 observations, too. In performing the Discriminant Analysis is good to check whether or not the covariance matrices $\Sigma_1 \, and \, \Sigma_2$ are equal. Result of Chi-square test convince us that they are significantly unequal, which implies that Quadratic Discriminant Analysis (QDA) is more prepared for the data.

Since PCs capture the majority information of the original data and at the same time showing the number of PCs are much less than that of original variables, it is easy to think of using the factors computed from PCA as the input futures for DA algorithm. The new features which are a linear combination of the original variables have several advantageous properties: (1) the interpretation of them allows us to detect patterns in the initial data space. (2) It is a very large reduced number of factors. Moreover, we can remove noise from the dataset by using only the most relevant factors.

4.4.1. Linear discriminant analysis

Although the covariance matrices from two classes are not equal, we run Linear Discriminant Analysis (LDA) to compare its performance when compared to the other

methods. Using the classification rule generated from 37 observations to re-classify them gives the results below. Table 4.5 and Table 4.6 respectively summarize the number of misclassification and the apparent error rate of misclassification. For instance, LDA method on original variables correctly classify 15 out of 18 as unsafe states, 3 states were misclassified which yield an apparent error rate of 16.7% and correctly classify 17 out of 19 as safe states also 2 states were misclassified causing an error rate of 10.5%, hence the percentage of correctness is 86.5%. However, the results of LDA method on PCs in Table 4.6 27.8% of the unsafe States and 36.8% of the safe States are misclassified. We acquire 67.6% correct rate. Thus we conclude that LDA on original variables does perform better than LDA on PCs with an 18.9% increase in classification accuracy.

Table 4.6: Classification result of applying LDA on the Original Variables

		CATEGORY	Predicted Group Membership		Total
			safe	***unsafe***	
Original	Count	***safe***	17	2	19
		unsafe	3	15	18
	%	***safe***	89.5	10.5	100.0
		unsafe	16.7	83.3	100.0

86.5% of original grouped cases correctly classified.

Table 4.7: Classification result of applying LDA on the PCs

		CATEGORY	Predicted Group Membership		Total
			safe	***unsafe***	
Original	Count	***safe***	12	7	19
		unsafe	5	13	18
	%	***safe***	63.2	36.8	100.0
		unsafe	27.8	72.2	100.0

67.6% of original grouped cases correctly classified.

4.4.2. Quadratic discriminant analysis

Let's have a look at results from QDA method on original variables. Table 4.6, in which 19 out of 19 unsafe states are correctly classified, its apparent error rate is 0% and no

misclassifications of 18 safe states. The rate of correct classification is consequently 100%, better than the value (86.5%) of using LDA. Again, review the number and the proportion of misclassification table for QDA on PCs case which showed in Table 4.7. The rate of correct classification is 75.76%.

Table 4.8: Classification result of applying QDA on the original Variables

		CATEGORY	Predicted Group Membership		Total
			safe	***unsafe***	
Original	Count	***safe***	19	0	19
		unsafe	0	18	18
	%	***safe***	100	0	100.0
		unsafe	0	100	100.0

100% of original grouped cases correctly classified.

Table 4.9: Classification result of applying QDA on the PCs

		CATEGORY	Predicted Group Membership		Total
			safe	***unsafe***	
Original	Count	***safe***	14	5	19
		unsafe	4	14	18

	%	***safe***	73.7	26.3	100.0
		unsafe	22.2	77.8	100.0

75.7% of original grouped cases correctly classified.

4.4.3. K-nearest neighbour

The result of KNN for the k=3 on both original variables and PCs are shown in Table 4.8 and Table 4.9 respectively. The result shown that on original variables records 30.12% misclassification rates and 83.5% correct classification rate when run on PCs

Table 4.10: Classification result of applying KNN for k=3 on the original Variables

			Predicted Group Membership		
		CATEGORY	***safe***	***unsafe***	Total
Original	Count	***safe***	16	3	19
		unsafe	8	10	18
	%	***safe***	84.21	15.79	100.0
		unsafe	44.44	55.56	100.0

69.9% of original grouped cases correctly classified.

Table 4.11: Classification result of applying KNN for k=3 on the PCs

		CATEGORY	Predicted Group Membership		Total
			safe	***unsafe***	
Original	Count	***safe***	18	1	19
		unsafe	5	13	18
	%	***safe***	94.73	5.26	100.0
		unsafe	27.78	72.22	100.0

83.5% of original grouped cases correctly classified.

4.4.4. Classification trees

The CT analysis classification result summary shown in table 4.10 and table 4.11 below, in which correctly classified 59.5% of the observations using original variables and 73.0% using PCs.

Table 4.12: Classification result of applying CT on the original Variables

		CATEGORY	Predicted Group Membership		Total
			safe	***unsafe***	
Original	Count	***safe***	17	2	19
		unsafe	13	5	18
	%	***safe***	89.5	10.5	100.0
		unsafe	72.2	27.8	100.0

59.5% of original grouped cases correctly classified.

Table 4.13: Classification result of applying CT on the PCs

			Predicted Group Membership		
		CATEGORY	***safe***	***unsafe***	Total
Original	Count	***safe***	14	5	19
		unsafe	5	13	18
	%	***safe***	73.7	26.3	100.0
		unsafe	27.8	72.2	100.0

73.0% of original grouped cases correctly classified.

4.4.5. Logistic discriminant analysis

We now apply LgDA to see how well the logistic model can discriminate between the groups, the results shown in Table 12 and Table 13, that LgDA on original variable has classified the observation more accurately compared with using PCs, with correct classification rate of 81.0% a 10.8% increase in classification accuracy.

Table 4.14: Classification result of applying LgDA on the original Variables

	Predicted Group Membership		
CATEGORY	***safe***	***unsafe***	Total

Original	Count	***safe***	16	3	19
		unsafe	4	14	18
	%	***safe***	84.21	15.79	100.0
		unsafe	22.22	77.78	100.0

81.0% of original grouped cases correctly classified.

Table 4.15: Classification result of applying LgDA on the PCs

			Predicted Group Membership		
		CATEGORY	***safe***	***unsafe***	Total
Original	Count	***safe***	14	5	19
		unsafe	6	12	18
	%	***safe***	73.68	26.32	100.0
		unsafe	33.33	66.67	100.0

70.2% of original grouped cases correctly classified.

4.5 Discussion

Though each of the applied discriminant techniques attempts to separate groups by different methods, they all found many of the same variables as significant predictors of group membership. Table 4.15 gives the summary of the rate of type I error (misclassify 'safe' to 'unsafe), rate of type II error (misclassify 'unsafe' to 'safe') and rate of correct classification for each method.

Table 4.16: summary of performance of all five methods

Procedures	***Sensitivity***	***Specificity***	***Correct classification***	***Error rate***
LDA on original variables	0.89	0.83	0.865	0.135
LDA on PCs	0.63	0.72	0.676	0.324
QDA on original variables	1.00	1.00	1.000	0.000
QDA on PCs	0.74	0. 78	0.757	0.243
KNN on original variables for k=3	0.84	0.56	0.699	0.301
KNN on PCs for k=3	0.94	0.72	0.835	0.165
CT on original variables	0.89	0.28	0.595	0.405
CT on PCs	0.74	0.72	0.730	0.270
LgDA on original variables	0.84	0.78	0.810	0.190
LgDA on PCs	0.74	0.67	0.702	0.298

The overall classification rates for each method are displayed in Table 4.16. The QDA achieved the greatest overall classification accuracy, correctly classifying 100% of the states using original variables and 75.7% using PCs. Therefore, QDA will be used when classifying State without using actual crime rates.

LDA and QDA both had consistent estimates of classification performance. However, LDA outperformed QDA by about 13% of the classification rate using original variables. QDA's classification rate is also greater for PCs. Therefore, QDA should be the preferred method for classification when deciding between these two.

Researchers should be cautious when using LDA due to the violations of its assumptions, which may complicate interpretations of significance (Manly, 2005) and call into question whether the models can be generalized to out of sample observations (Ousley and Jantz, 2012). It should also be recognized that LDA does not perform optimally under conditions violating its assumptions (Huberty and Olejnik, 2006). While QDA's assumptions were better upheld, caution is also warranted when applying it. Holden and Kelley (2010) note QDA's increased sensitivity to misclassifications and errors in the training data. Furthermore, QDA is known for developing over-fit models that do not perform as well on out of sample data. Thus, models developed on available data may be optimistic when applied to new instances.

KNN does not construct functions for classification, it classifies by observing group membership of nearby observations. Furthermore, Beyer *et al.* (1999) reported that smaller values of *k* are more sensitive to dimensionality, and a small value of *k* (3) was used. However, Table 4.15 shows that the large difference between in classification rate between the KNN and QDA with about 30%.

While KNN is more appropriate in situations where normality assumptions are violated, it may not be a very useful method in the context of crime analysis. Becouse, it is a nonparametric statistic that is good for situations where data are not normally distributed. In this analysis the data for many of the variable subsets ended up approximating normality. In some cases where normality was violated variables transformed to normal. Another issue with KNN is it assumes homogeneity of variance. Every evaluation of covariance matrix equality resulted in significance regardless of how many or few variables were included. It is difficult for this assumption to be met, but it may be easier to satisfy in an analysis with fewer groups. Furthermore, it is difficult to evaluate KNN assumptions because of the distributional situation for which it is most appropriate – equal covariance matrices and lack of multivariate normality; tests for equality of covariance matrices are known for their sensitivity to violations of their multivariate normal assumptions (Rencher, 2002). Tests for homogeneity of variance that are robust to departures from normality have been proposed (O'Brien, 1992), but they are neither common nor built into any available software known by the author.

While CT did not perform as well as LDA its performance was not far off, based upon the available classification estimates in Table 4.16. Furthermore, while it does have reduced performance, it has many advantages that simplify analysis and interpretation as well as reducing uncertainty of the results. CT makes no assumptions pertaining to data distributions. Thus, assumptions need not be tested and there are no violations that can invalidate results, compromise performance, or complicate interpretation. Furthermore, these conditions remain consistent regardless of which and how many groups and variables are included. CT, by nature of its workings, has built-in variable selection and is robust to outliers, errors, and misclassifications in the data.

LDA may be a better choice, but with missing observation CT will performed outstandingly, whereas commercial Discriminant Analysis program do not predict classification for cases with incompletely measured predictor variables. The most frequently used implementation of LDA (i.e. SAS, SPSS and SYSTAT) are designed only for complete suites of measurement on each individual, and delete any case missing one or more predictor variables.

To overcome the above, a few researchers have used multiple regression to predict missing values, or have substituted means for missing values (as in SPSS). These approaches were discredited by Schafer (1997) and Schimert et al. (2000) who noted that deleting cases throw away important information.

LgDA performance as shown in Table.4.15 is appreciative which correctly classified 81% and 70% using original variables and PCs respectively. Is therefore, recommended to be used instead of QDA when normality assumption is violated.

CHAPTER FIVE

SUMMARY AND CONCLUSION

5.1 Introduction

This thesis mainly studies the roles of the Principal Component Analysis and Discriminant Analysis in the pattern recognition and a comparative evaluation of five statistical discriminant analysis techniques in crime data analysis: linear discriminant analysis (LDA); quadratic discriminant analysis (QDA); k-nearest neighbour analysis (KNN); classification trees (CT); and Logistic discriminant analysis. Performance evaluations were based upon correct classification rates for each method. Data consisted of 13 variables affecting crime rates from 37states of Nigeria were used. It reveals effectiveness of these two primary techniques in excavating internal structure of data with large dimensionality, but suggests also that problems exist and more improvements could be made to enlarge the significance and accuracy of such pattern recognition process.

5.2 Conclusion

Principal Component Analysis has been successfully applied into our data by extracting four PCs out of the 13 original variables, which implies a great dimensionality reduction. In addition, these PCs account for almost 75.024% variance of the original dataset, thus 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. Illiteracy rates, population density and GDP collaboratively make contribution to PC1. PC2, PC3, and PC4; represent population factor, political factor and factors related to gender respectively. That

is to say, the above six variables can considerably help when classifying an area as 'safe' or 'unsafe' crime wise.

Overall, QDA performed the best, indicated by the classification rate estimates in Table 4.15. While the classification rate for the LDA is better than that of CT on original variables. LgDA and KNN analysis did perform well in which correctly classified 80% and 77% of the States.

In their discussion of LDA and QDAs' assumptions, Ousley and Jantz (2012:317) state, "If groups show very different levels of variability other statistical procedure may be necessary, although they may not produce the most accurate statistics." The results of this analysis support their statement; the nonparametric procedures evaluated here did not perform as well as QDA and LDA.

If the conditions are such that a nonparametric method should be applied, then CT is recommended. QDA is advantageous in that it does not assume variance homogeneity. However, it is sensitive to errors in the training data and is known for developing over fit models that do not perform well when applied to data not used during model estimation. KNN is not recommended because it assumes variance homogeneity and the KNN models in this analysis became over-fit very quickly. LDgA on the other hand assume the likelihood of the group to have exponential form. CT has many advantages among which are easy comprehension, interpretation, and application. However, users must be careful when applying them to groups with very different sample sizes.

The CT accuracy in this analysis was not far below that of LDA. Further work with CTs is recommended). CT does not make any assumptions about the data that if violated can confound interpretation or call into question the validity of any statistics or estimates made. Thus, there is no need for additional statistical tests of assumptions either. There is also no need for a variable selection analysis because variable selection is inherent to CT algorithms. CTs can incorporate continuous, categorical, or both variable types in its analysis. They can include observations with missing values and have a variety of methods for handling missingness. Lastly, CTs are more easily applied and interpreted than other discriminant analysis methods.

The results of this analysis indicate that LDA's assumptions are difficult to satisfy with crime data, especially homogeneity of variance. If tests for equality of variance are significant and a nonparametric statistic is required, then CT is recommended. Its performance was similar to that of QDA. LgDA is also recommended as does not make any distributional assumption. KNN is not recommended due to its erratic and greatly differing classification rates. Furthermore, its assumptions are violated by these data and KNN models rapidly became over-fit in this analysis. While QDA's assumptions were satisfied by the data and its classification rates have greatest accuracy, it is also recommended for use in classifying states with as high and low crime rate without using actual crime statistics.

To make up the fact that the common features PCA extract is the "global" features, using not enough or inadequate representatives for discriminant on class from another, a more suitable method for discriminating different pattern classes, Discriminant Analysis could then be applied to classify the states which have no data records of crime commission

rates. Apply DA on original variables and PCs' space respectively. For our case, model of discriminant function on the original variables performs best.

Safety is the most important issue for everyone in Nigeria, let alone everyone in the world. We now see that we, as Nigerians, are willing to give up some of our freedom to make sure that our children grow up in a safe environment. With the help of the multivariate statistical methods of Principal Component Analysis and Discriminant Analysis we have created a method that helps to classify a State as safe or unsafe without using the actual crime rate statistics of the State. We have shown certain variables such as population, sex ratio, voter turnout, police strength, unemployment rates, among others are suitable for analysis to classify a State as safe or unsafe and to help improve the safety of the State. The State Governments will be able to decrease the crime rates of their States by increasing or decreasing certain values of our variables.

5.2 Recommendation

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

CT is recommended either as replacement for LDA or as a supplement whenever data do not meet relevant assumptions. CT is also highly recommended when data contain important cases with missing predictor variables. LgDA is non parametric discriminant method recommended to be used in place of QDA when data distribution is not normal. KNN is not recommended because it assumes variance homogeneity and also becomes highly unreliable in higher dimension.

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

For effective crime prevention and control, and promotion of safety and security through effective law enforcement, intelligence gathering and analysis, police criminal statistics need to be useful. In view of this the Nigeria police force policy makers should seriously consider the development of non-uniform and non-policing personnel who are professional in Criminology and Social Statistics to handle (collation and analysis) crime data and suggest lines of research and planning. With at least one of such qualified professional in each state police command and about five in force headquarters Abuja, the problem of crime statistics should be solved.

5.3 Improvements for future study

(1) The variables we choose is based on the historically study of possible factors causing crime, and also are up to the accessibility of the variables. If data resources are sufficient, we can analyze additional variables especially category variables such as major traffic mode, alcohol assumption amount because discriminant analysis has superiority when dataset involves category variables which are not verified in our research.

(2) Discriminant analysis can only tell the likely class of a certain observation but do not know certain properties of an observation in a very specific degree such as within a confidence interval. Hence, we may further use regression method to find the significant variables for further prediction work.

References

Abramson, D. (1944), *Crime and Human Mind*, New York: Columbia University Press

Acuña E, and Rodriguez C. (2004). The Treatment of Missing Values and its Effect on Classifier Accuracy. *Classification, Clustering, and Data Mining Applications*:639-647.

Adeyemi, A. A. (1993) "Information Management for National Development, Planning and Security" in T. N. Tamuno; I. L. Bashir; E. E. O. Alemika and A. O. Akano eds. *Policing Nigeria : Past, Present and Future* (Lagos : Malthouse Press Limited).

Adeyemi, M. N. (1989) "Information Management for Crime Control in Nigeria" in B. J. Takaya ed. *Security Administration and Human rights in Nigeria,* Jos: University of Jos, Centre for Development Studies and the Federal Ministry of Internal Affairs

Al-Ghazali (2004), Daily Trust. 13/7/2004, pp. 40.

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

Ackerman, W.V. and Murray A.T. (2004), Assessing spatial patterns of crime in Lima, Ohio, USA Cities. *Published by Elsevier Ltd. Printed in Great Britain*. 21(5):423–437, 2004

Ayoola, A.O. (2008). causes of corruption. *This Day,* 20/2/2008

Azubuike, C. (2007), Crime and control strategies. *Daily Sun,* Tues. 10/2/2009.

Beyer K, Goldstein J, Ramakrishnan R, and Shaft U. (1999). When is "nearest neighbor" meaningful? 1540:217-235.

Breiman L, Friedman JH, Olshen RA, and Stone CL. 1984. *Classification and regression trees*. Belmont, CA: Wadsworth. International Group, Inc.

Birkby WH. 1966. An evaluation of race and sex identification from cranial measurments. *American Journal of Physical Anthropology* **24(2):**21-28.

Brace CL. 1995. Region does not mean "race"—reality versus convention in forensic anthropology. *Journal of Forensic Sciences* **40(2):**171-175.

Cattell, R. B. (1966), The Scree test for the number of factors. *Multivariate Behavioral Research,* **2(1):**245-276.

Carvell, I. G. and Swinfen, G. E. (1970), *Criminal law and Procedure*. London: Sweet and Maxwell.

Cleen (1993), http://www.cleen.org/crime%20survey1993.pdf

Cleen (2003), hpp://www.cleen.org/crime%20survey2003.pdf

Cleen (2004), http://www.cleen.org/crime%20survey%202004.pdf

Cleen (2005), http://www.cleen.org/crime%20survey%202005.pdf

Cleen (2006), http://www.cleen.org/crime%20survey%202006.pdf

Cleen (2007), The Nigeria Police as at November 2007. http://www.cleen.org/summary%20of%20crime%20statistics%20in%20nigeria%202007.p

Connally M. (2004). Identifying covariance deifferences in comparisons of linear versus quadratic classification rule [dissertation]. Athens: University of Georgia.

Corcoran, J. J., Wilson, I. D. and Ware, J. A. (2003), Predicting the geo-temporal variations of Crime and disorder. *International Journal of Forecasting.* **19(4)**: 623 – 634

Cover T. M. and Hart P. E. (1967). Nearest neighbour Pattern classification. *IEEE Transaction on Information Theory*. **13** 21-27.

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

Dauda, M (2007). A chronicle of Crime Data Using Some Multivariate Techniques. Unpublished Msc. Thesis. Ahmadu Bello University, Zaria Nigeria

Ding Y, and Simonoff JS. (2010). An investigation of missing data methods for classification trees applied to binary response data. **11**:131-170.

Donald F. Morrison, (2005) Multivariate Statistical Methods (4th Edition), Brooks/Cole, a division of Thomson Learning, Inc. (2005)

Duran, B. S., and Odell, P. L. (1974), Cluster Analysis: a Survey, *Lecture Notes in Economics and Mathematical Systems*, New York: Springer-Verlag.

Ebbe, Obi N. I. (undated), World Factbook of Criminal Justice. *www.ojp.gov/bjs/pub/ascii/wfbcjnig.txt -54k.*

Elizabeth, C. M. (2004), Geographies of Urban Crime: An Intraurban Study of Crime in Nashville, TN; Portland, OR; and Tucson, AZ. PhD Dissertation, University of Arizona.

Enas, G. G., and S. C. Choi. (1986). Choice of the Smoothing Parameter and Efficiency of K-Nearest Neighbor Classification. *Computers and Mathematics with Applications* 12A: 235–244.

Fang, Y (2011). Multivariate Method Analysis of Crime Data in Los Angeles Communities; Published Msc. Thesis. University of Califonia, Los Angeles

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

Fisher RA. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics* **7(2):**179-188.

Fisher TD, and Gill GW. 1990. Application of the giles-elliot discriminant function formulae to a cranial sample of northwestern plains indians. In: Gill GW, and Rhine S, (Ed). *Skeletal attribution of race.* Albuquerque, NM: Maxwell Museum of Anthropology.

Fix E. And Hodges J. L. (1951). Discriminatory Analysis. Nonparametric Discrimination: consistency properties. Report No. 4, project No. 21-29-2004. Brooks Air Force Base. USAF School of Aviation Medicine.

Feldman, M. P. (1997), *The Psychology of Crime*. Cambridge: University Press.

Fielding AH. 2007. *Cluster and classification techniques for the biosciences*. New York: Cambridge University Press.

Feldesman M. (2002). Classification trees as an alternative to linear discriminant analysis. *American Journal of Physical Anthropology* **11(9):**257-275.

Gabor, T. (1978), Crime displacement: the literature and strategies for its investigation. *Crime and Justice* **6(7):**100-117.

Giles E, and Elliot O. 1962. Race identification from cranial measurements. *Journal of Forensic*

Sciences **7(2)**:147-157.

Greenburg, S. And W. Rohe (1984): Neighbourhood Design and Crime. *Journal of the American Planning Association.* **50**:48-61

Grubesic, T. H. (2006),: On the Application of Fuzzy Clustering for Crime Hot Spot Detection. *Journal of Quantitative Criminology. Springer Netherlands.* **22(1).**

Gulumbe, S. U.,Dikko H. G. And Bello Y. (2012): Analysis of Crime Data Using Principal Component Analysis: A case study of Katsina State. *CBN Journal of Applied Statistics.* **3(2):**39-48

Hardle, W. and Simar, L. (2003): *Applied Multivariate Statistical Analysis*, 2nd edition. New York: Springer-Verlag.

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

Holden JE, and Kelley K. 2010. The effects of initially misclassified data on the effectiveness of

discriminant function analysis and finite mixture modeling. *Educational and Psychological Measurement* ***70(1):***36-55.

Hooton, E. (1939), *The American Criminal*, Cambridge, MA.: Harvard University press,;

Howells WW. (1989). Skull shapes and the map. Papers of the Peabody Museum of Archaeology

and Ethnology. Cambridge, MA. p 1-189.

Huberty C. J. (1975). The Stability of three Indices of Relative variable contribution in Discriminant Analysis. *Journal of Experiment Education*. **44**, 59-64

Huberty CL, and Olejnik S. (2006). *Applied manova and discriminant analysis*. Hoboken, NJ:

John Wiley & Sons.

Ikiro, M. (2007), State of the Nigeria Police Force as at November 2007. http://nigeriapolice.org/public/modules/mastop_publish/?tac=STATE_OF_NIGERIA_POLI

Johnson, C. P. (2000), Crime Mapping and Analysis Using GIS. Geomatics Conference on Geomatics in Electronic Governance, January 2000, Pune Geomatics Group, C-DAC, Pune University Campus, Pune 411007.

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

Kaiser H. F. (1970). A second Generation Little Jiffy. Psychometrika. **35** 401-415.

Karofi, U. A. (2005). Drug and Criminal Behaviour in Penang, Malaysia: A Multivariate Analysis. *Bangladesh e-Journal of Sociology.* **2(2)** 1-26.

Khattree R, and Naik DK. (2000). *Multivariate data reduction and discrimination with SAS software.* New York: Wiley.

Kpedekpo, G.M.C. and Arya, P.L. (1981), *Social and Economics Statistics for Africa*. London: George Allen and Unwin.

Kutigi, I. (2008), Puberty, ignorance and evil intentions as strong crime factors. *Tribune*, Mond. 23/06/2008.

Lance, G. N., and Williams, W. T. (1967), A General Theory of Classificatory Sorting Strategies: I. Hierarchical Systems. *Computer Journal,* **9,** 373–380.

Linderman, R. H., Peter, F. and Gold, R. Z. (1980), Introduction to Bivariate and Multivariate Analysis. Scott, Foresman and Company.

Loftsgaarden, D. O., and C. P. Quesenberry. 1965. A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics* **36**: 1049–1051.

Louis, S., Cookie, W. S., Louis, A. Z. and Sheldon, R. E. (1981), *Human Response to Social Problems.* Illinois: The Dorsey Press.

Lombroso, C. (1911), *The Criminal Man*, New York: Putman.

Manly BFJ. (2005). *Multivariate statistical methods*: A primer. New York, NY: Chapman & Hall/CRC.

McGuffog, I., Western, J.S. and Mullins, P.(2001), Exploratory Spatial Data Analysis Techniques for Examining Urban Crime. *British Journal of criminology,* **41(2):**309 – 329.

McLachlan, G. J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley.

Mercer, D. A. (2013). Nonparametric Discriminant Analysis in Forensic Ancestry Estimation: An Assessment of Utilitzed and Alternative Statistical Methods." Master's Thesis, University of Tennessee, 2013. http://trace.tennessee.edu/utk_gradthes/1644

Muga, E. (1975), Crime and Delinquency in Kenya. Nairobi: East African Literature Bureau.

Nat.Sec. (2005), http://www.nigeriadailynews.com/nigeria/news_item.asp?NewsID

=12. Tuesday, March 15/2005.

Nigeria (1991), Nigeria Incidence of Crime Rate. http://www.photius.com/countries/nigeria_security/nigeria_nationalsecurity_ir

NPC (2007), Federal Republic of Nigeria Official Gazette, 24(94).2

Nzeakah, G. (1978), Magistrate raps police. *New Nigeria*. 15th July, 1978.

Okereke, O. G. (1993), Public Attitudes towards the Police Force in Nigeria, *Police Studies.* 16(3):113-121.

Ousley S, Jantz R, and Freid D. (2009). Understanding race and human variation: Why forensic

anthropologists are good at identifying race. *American Journal of Physical Anthropology*

13(9):68-76.

Oyebanji, J. O. (1982), Economic Development and the Geography of Crime: an Empirical Analysis. *GeoJournal. Springer Netherlands.6 (5)*

Printcom (2003), http://support.sas.com/onlinedoc/912/getDoc/common.hlp/images/

copyrite.htm.

Radoff, D. (1993), GIS Responds to Emergency Management. *International GIS Sourcebook, pp. 209-210.*

Radzinowicz, L. and King, L. (1977), The Growth of Crime: The International Experience, London: Hamish Hamilton, pp. 15.

Rencher A. C. (2000). *Linear models in Statistics*. New York: Wiley

Rencher A. C. (2002). *Methods of multivariate analysis*. Hoboken, NJ: John Wiley & Sons.

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

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

Sam, O. (2003). Crime war rages in Nigeria. *BBC News*. Thursday, 20 February, 2003, 10:14 GMT.

Savoie, J. (2008), Analysis of the Spatial Distribution of Crime in Canada: Summary of Major Trends. *Crime and Justice Research Paper Series. Statistics Canada Catalogue* **85**:561-615.

Sauer and Wankmiller J. C. (2009). The assessment of ancestry and the cocept of race. InL: Blan S. And Ubelaka D. H. Editors. Handbook of forensic anthropology and archaeology. Walant creek, CA: left coast pree. 187-200.

Sellin, T. and Wolfgang, M.E. (1964). 'Measurement of Delinquency', in New Thinking in Criminal Statistics. *Journal of Criminal Law, Criminology and Police Science* **56(3)**: 227-284.

Shepherd, J. (1981), Sociology' St. Paul, Minn: West Publishing.

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

Soren, H. (2006), Example of multivariate analysis in R – Principal component analysis (PCA). httpgene://tics.agrsci.dk/statistics/courses/Rcourse-Djf2006 / day3/ PCA-notes

StatSoft (2004), http://www.statsoft.com/textbook/stclvan.html

Sutherland, E.H. (1939), The White-Collar Criminal. *American Sociological Review*, **5.** 1-12.

Tappan P. W. (1960): Crime, Justice and correction. New York Mc Graw-Hill

Taylor L. (1984), Born Criminal: The Genetic causes of Criminal Behaviour, London: Greenwood Press.

Uche, O. (2008), Nigeria Prison Robbed by Criminals. http://www.whichwayNigeria.Net/Nigerian-prison-robbed-criminals/

Udo, R. K. and Mamman, A. B. (1993), Nigeria: Giant in the Tropics. Lagos: Gabumo Publication Company.

UN (1995), States in Disarray: The Social Effects of Globalisation, UN Research Institute for Social Development.

Usman, U., Yakubu, M. And Bello, A. Z. (2012): An Investigation on the rate of crime in Sokoto State Using Principal Component Analysis. *Nigerian Journal of Basic and Applied Sciences,* **20(12)***:152-160*

Ward, J. H. (1963), Hierarchical grouping methods to optimize an objective function. *Journal of the American Statistical Association* **58**, 236-244.

WBE (1974), The World Book Encyclopaedia. Vol.4. USA. Field Enterprises Educational Corporation. pp. 908 – 912

Williams FLE, Belcher RL, and Armelagos GJ. (2005). Forensic misclassification of ancient nubian crania: Implications for assumptions about human variation. *Current Anthropology* **46(2)**:340-346.

Wilson, P. (2009), Nigeria: Crime Rate Drops in Lagos*. Daily Independent*, 12/ 01/2009.

APPENDIX A

TABLES NOT LISTED IN THE CHEPTER

State	pop	DENS	SEX	ILLT	MOD	PVR	UNR	VOT	POL	GDP	AVRG	ARRT	SAI
Abia	2881380.00	580.40	102.50	11.30	30.30	57.40	11.20	77.90	185.60	8687442705	5.2	160	9.4
Adamawa	3178950.00	81.90	102.80	39.70	39.20	74.20	33.80	50.00	173.39	4582045246	6.2	348	6
Akwa Ibom	3902051.00	576.20	109.00	16.10	43.30	53.00	24.50	76.20	163.50	11179887963	5.5	262	5
Anambra	4177828.00	867.50	108.30	7.90	45.00	56.00	12.20	57.50	177.82	6764219562	5.4	361	0.5
Bauchi	4653066.00	93.20	107.80	53.30	41.90	73.00	41.40	63.80	123.06	4713858180	7.1	130	6.5
Bayelsa	1704515.00	181.60	112.70	19.50	28.60	47.00	23.90	85.60	195.19	4337065923	5.4	198	0.8
Benue	4223641.00	136.00	105.30	39.50	32.70	67.10	14.20	43.50	158.58	6864209262	6.2	132	8.7
Borno	4171104.00	55.30	108.60	58.50	51.30	55.10	29.10	49.50	151.21	5175165142	6.3	112	11
C/River	2892988.00	133.70	106.90	20.30	26.00	52.50	22.50	63.20	302.94	9292059207	6	195	1.2
Delta	4112445.00	238.60	102.50	18.10	36.20	63.30	27.30	69.40	218.75	16749250544	5	185	145
Ebonyi	2176947.00	339.00	91.90	21.00	15.20	73.60	23.10	47.90	104.23	2732472739	5.8	142	2.5
Edo	3233366.00	163.10	104.00	21.60	59.10	65.60	35.20	37.50	258.77	11888446884	5.5	240	169
Ekiti	2398957.00	407.40	103.60	25.00	51.10	52.40	12.10	34.20	178.54	2848372512	4.7	211	18
Enugu	3267837.00	426.60	99.50	24.20	42.20	62.50	25.20	62.50	124.24	4396590769	5.1	133	5
Gombe	2365040.00	131.50	109.60	54.50	46.30	74.20	38.70	58.40	194.37	2500467306	7	91	7.9
Imo	3927563.00	757.80	106.80	6.50	30.40	50.50	26.10	83.60	121.57	14121670484	5.6	147	2.5
Jigawa	4361002.00	177.90	103.90	54.50	56.00	74.10	35.90	56.60	112.22	2988014405	6.8	261	0.4
Kaduna	6113503.00	133.70	105.30	33.00	38.20	61.50	30.30	65.80	186.42	10334763785	6.9	214	10
Kano	9401288.00	444.90	106.70	37.40	40.80	65.60	21.30	53.20	104.36	12393103864	6.6	1281	3.9
Katsina	5801584.00	232.30	105.90	53.70	46.20	74.50	28.10	52.40	104.49	6022655197	6.5	350	1.7
Kebbi	3256541.00	86.30	99.80	63.40	46.30	72.00	25.30	56.40	142.33	3290847166	5.5	144	2.1
Kogi	3314043.00	112.00	106.60	25.20	58.70	67.10	14.40	42.70	128.27	4642794262	5.4	102	43
Kwara	2365353.00	68.60	106.10	27.40	57.00	61.80	7.10	36.00	233.03	3841827534	5.4	95	12
Lagos	9113605.00	2607.00	107.90	12.60	66.20	48.60	8.30	31.80	174.84	33679258023	4.9	382	27
Nasarawa	1869377.00	68.50	103.00	32.50	60.00	66.40	36.50	50.00	384.89	3022828885	6.4	106	2.6
Niger	3954772.00	53.40	106.00	52.60	59.60	33.80	39.40	46.80	275.39	6002007080	5.8	188	3.9
Ogun	3751140.00	222.90	98.20	22.70	69.20	62.30	22.90	28.00	197.86	10470415017	4.4	188	18
Ondo	3460877.00	227.80	104.20	12.60	44.20	45.70	12.50	30.10	259.62	8414302623	4.9	132	402
Osun	3416959.00	392.80	103.40	12.40	65.30	37.90	3.00	39.60	178.08	7280597521	5.1	196	16
Oyo	5580894.00	197.60	101.00	16.20	49.50	51.80	8.90	33.60	187.57	16121670484	5.1	230	67

Plateau	3206531.00	117.80	100.50	41.40	55.90	74.10	25.30	62.50	253.20	5154059937	6.6	163	1.6
Rivers	5198716.00	498.30	109.50	8.60	33.90	50.40	25.50	76.30	209.80	21073410422	5.6	349	1.6
Sokoto	3702676.00	109.60	102.60	63.60	59.40	81.10	17.90	40.10	125.69	4818615261	6.2	212	15
Taraba	2294800.00	38.10	109.00	54.60	29.80	68.40	12.70	55.30	220.24	3397790217	6.4	114	4.1
Yobe	2321339.00	49.50	108.10	65.00	49.70	62.50	28.50	45.30	126.74	2011499081	6.6	257	2.7
Zamfara	3278873.00	93.20	100.10	68.50	43.60	70.80	42.60	51.70	118.73	4123829498	5.5	75	3
FCT Abuja	1406239.00	181.40	111.40	30.20	55.60	43.50	25.50	42.20	857.61	5010968012	5.8	258	15

Table A.1: dataset before transformation

State	Lgpop	LgDens	Sex	Avrg	ln-Iltr	Mod	Pvr	Unr	Vot	SqrLgPol	LnLnLgGDP	LgArrst	LgSAI
ABS	6.459601	2.763727	102.5	5.2	2.424803	30.3	57.4	11.2	77.9	1.50618	0.831367	2.20412	0.972203
ADS	6.502284	1.913284	102.8	6.2	3.681351	39.2	74.2	33.8	50	1.496337	0.818944	2.541579	0.781037
AKS	6.591293	2.760573	109	5.5	2.778819	43.3	53.5	14	76.2	1.48779	0.836129	2.418301	0.69897
ANS	6.620951	2.938269	108.3	5.4	2.066863	45	56	12.2	57.5	1.499994	0.826568	2.557507	-0.26761
BUS	6.667739	1.969416	107.8	7.1	3.975936	41.9	73	41.4	63.8	1.445724	0.819506	2.113943	0.811575
BYS	6.231601	2.259116	112.7	5.4	2.322388	28.6	47	23.9	85.6	1.513426	0.817853	2.296665	-0.08092
BES	6.625687	2.133539	105.3	6.2	3.676301	32.7	67.1	14.2	43.5	1.483323	0.826851	2.120574	0.93902
BNS	6.620251	1.742725	108.6	6.3	4.069027	51.3	55.1	29.1	49.5	1.47634	0.821347	2.049218	1.054996

CRS	6.461347	2.126131	106.9	6	3.010621	26	54.2	28.2	63.2	1.575232	0.832644	2.290035	0.08636
DTS	6.6141	2.37767	102.5	5	2.895912	36.2	63.3	27.3	69.4	1.529689	0.843607	2.267172	2.161368
EBS	6.337848	2.5302	91.9	5.8	2.322388	15.2	73.6	23.1	47.9	1.420561	0.808523	2.152288	0.404834
EDS	6.509655	2.212454	104	5.5	3.072693	59.1	65.6	35.2	37.5	1.553356	0.837278	2.380211	2.228683
EKS	6.380022	2.610021	103.6	4.7	2.493205	51.1	52.4	12.1	34.2	1.500578	0.809374	2.324282	1.249443
ENS	6.51426	2.630021	99.5	5.1	3.186353	42.2	62.5	25.2	62.5	1.447156	0.818124	2.123852	0.696356
GBS	6.373838	2.118926	109.6	7	3.883624	46.3	74.2	38.7	58.4	1.512822	0.806699	1.959041	0.896526
IMS	6.594123	2.879555	106.8	5.6	1.871802	30.4	50.5	26.1	83.6	1.443893	0.840473	2.167317	0.390935
JGS	6.639586	2.250176	103.9	6.8	3.998201	56	74.1	35.9	56.6	1.431807	0.810351	2.416641	-0.37675
KDS	6.78629	2.126131	105.3	6.9	3.496508	38.2	61.5	30.3	65.8	1.506815	0.834653	2.330414	0.998695
KNS	6.973187	2.648262	106.7	6.6	3.621671	40.8	65.6	21.3	53.2	1.420751	0.838052	3.107549	0.585461
KTS	6.763547	2.366049	105.9	6.5	3.983413	46.2	74.5	28.1	52.4	1.420941	0.824315	2.544068	0.240549
KBS	6.512757	1.936011	99.8	5.5	4.149464	46.3	72	25.3	56.4	1.467411	0.812312	2.158362	0.311754
KGS	6.520358	2.049218	106.6	5.4	3.226844	58.7	67.1	14.4	42.7	1.451938	0.819205	2.0086	1.63367
KWS	6.373896	1.836324	106.1	5.4	3.310543	57	61.8	7.1	36	1.53864	0.815432	1.977724	1.080987
LGS	6.95969	3.416141	107.9	4.9	2.533697	66.2	48.6	8.3	31.8	1.497545	0.856107	2.582063	1.435367
NAS	6.271697	1.835691	103	6.4	3.511545	60	66.4	36.5	50	1.607898	0.810587	2.025306	0.418301
NGS	6.597121	1.727541	106	5.8	3.962716	59.6	33.8	39.4	46.8	1.562033	0.824248	2.274158	0.593286
OGS	6.574163	2.34811	98.2	4.4	3.122365	69.2	62.3	22.9	28	1.515374	0.834898	2.274158	1.260787
ODS	6.539186	2.357554	104.2	4.9	2.533697	44.2	45.7	12.5	30.1	1.553814	0.830758	2.120574	2.604496
OSS	6.53364	2.594171	103.4	5.1	2.517696	65.3	37.9	3	39.6	1.500205	0.827987	2.292256	1.207904
OYS	6.746704	2.295787	101	5.1	2.785011	49.5	51.8	8.9	33.6	1.507701	0.842908	2.361728	1.823865
PLS	6.506035	2.071145	100.5	6.6	3.723281	55.9	74.1	25.3	62.5	1.550311	0.821267	2.212188	0.190332
RVS	6.715896	2.697491	109.5	5.6	2.151762	33.9	50.4	25.5	76.3	1.523747	0.847774	2.542825	0.201397
SKS	6.568516	2.039811	102.6	6.2	4.152613	59.4	81.1	17.9	40.1	1.448896	0.81994	2.326336	1.180413
TBS	6.360745	1.580925	109	6.4	4.000034	29.8	68.4	12.7	55.3	1.530652	0.812959	2.056905	0.61595
YBS	6.365739	1.694605	108.1	6.6	4.174387	49.7	73.6	23.5	45.3	1.450143	0.802178	2.409933	0.432969
ZFS	6.515725	1.969416	100.1	5.5	4.067316	43.6	70.8	42.6	51.7	1.440333	0.816849	1.875061	0.481443
FCT	6.148059	2.258637	111.4	5.8	3.407842	55.6	43.8	15.3	42.2	1.712685	0.820712	2.41162	1.169086

Table A.2: dataset after transformation

State	Category	Crimerates
ABS	0	108.45
ADS	1	236.08
AKS	0	115.32
ANS	0	53.42
BUS	0	82.27
BYS	1	136.4
BES	0	66.03
BNS	0	112.51
CRS	1	206.98
DTS	1	211.84
EBS	1	136.2
EDS	1	153.56
EKS	0	34.77
ENS	0	68.15
GBS	0	88.5
IMS	1	168.17
JGS	0	63.54
KDS	0	33.83
KNS	1	124.66
KTS	0	21.18
KBS	0	49.29
KGS	0	23.45
KWS	1	126.24
LGS	1	204.34

NAS	1	148.87
NGS	0	111.92
OGS	1	130.39
ODS	1	126.59
OSS	1	173.66
OYS	1	280.92
PLS	1	156.49
RVS	1	153.69
SKS	0	75.38
TBS	1	127.16
YBS	0	97.44
ZFS	0	34.16
FCT	0	95.86

Table A3. Crime rates