

**CANONICAL CORRELATION ANALYSIS OF
AGGRAVATED ROBBERY AND POVERTY IN
LIMPOPO PROVINCE**

by

TANDANAI RWIZI

A Dissertation Submitted in Accordance with the Requirements for
the Degree of

MASTER OF SCIENCE

in the subject

STATISTICS

at the

UNIVERSITY OF SOUTH AFRICA

Supervisor: Prof. J.O OLAOMI

May 2015

Declaration

I declare that the dissertation entitled **Canonical Correlation Analysis of Aggravated Robbery and Poverty in Limpopo Province**, which I hereby submit, is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references. I further declare that the dissertation manuscript has not been submitted at any other institution for any academic work.

Declaration date:_____

Signed:_____

Name: T. RWIZI

Acknowledgements

I wish to convey my sincere gratitude to those who contributed to this research manuscript. The following are acknowledged:

- ❖ The research would have remained a dream had it not been the mentorship of my supervisor Professor J. O Olaomi, for his assistance, guidance and encouragement throughout the course of my study.
- ❖ I also appreciate the assistance of the University of South Africa where the research was carried out in terms of the SPSS program which was used in the analysis. Many thanks to Dr EM Rapoo (Chair of Department of Statistics) for the support given from the initial stages of the research up to its present form.
- ❖ It is my pleasure to thank the support offered by the South African Police Service (SAPS) from the national level to the provincial police stations where samples of offenders were collected.
- ❖ I hereby gratefully acknowledge Mr Ssekuma's profound assistance in the typing process with the Latex program where I needed his help most.
- ❖ I would also extend my gratitude to Dr KN Phalanndwa (Institutional Coordinator – Editing and Proofreading Unit – University of Venda) for proofreading my research script.

Financial Assistance

The support provided by the National Research Foundation(NRF) to this research in respect of the costs is hereby gratefully acknowledged. The NRF is not responsible for the opinions or conclusions that have been conveyed in this study.

Dedication

I dedicate this dissertation to my wife Loveness and children: Sharon, Joyline, Wadzanai and Wellington for allowing me to be out of their company at the time when they needed me most.

Abstract

The study was aimed at exploring the relationship between poverty and aggravated robbery in Limpopo Province. Sampled secondary data of aggravated robbery offenders, obtained from the South African Police (SAPS), Polokwane, was used in the analysis. From empirical researches on poverty and crime, there are some deductions that vulnerability to crime is increased by poverty. Poverty set was categorised by gender, employment status, marital status, race, age and educational attainment. Variables for aggravated robbery were house robbery, bank robbery, street/common robbery, carjacking, truck hijacking, cash-in-transit and business robbery. Canonical correlation analysis was used to make some inferences about the relationship of these two sets. The results revealed a significant positive correlation of 0.219 (p -value = 0.025) between poverty and aggravated robbery at five per cent significance level. Of the thirteen variables entered into the poverty-aggravated model, five emerged as statistically significant. These were gender, marital status, employment status, common robbery and business robbery.

Key words: Canonical correlation analysis, poverty, aggravated robbery, Limpopo province, poverty-aggravated robbery model

Contents

Declaration	i
Acknowledgements	ii
Financial Assistance	iii
Dedication	iv
Abstract	v
List of Tables	xi
List of Figures	xiii
1 Background and Justifications	1
1.1 Introduction	1
1.2 The Research Problem	3
1.3 Purpose of the Study	4
1.4 Significance of the Study	4
1.5 Objectives of the Research	5
1.6 Research Questions	5
1.7 The Research Hypothesis	6
1.8 Overview of Methodology	6
1.9 Assumptions of the Study	6
1.10 Limitations and Delimitations	7
1.11 Abbreviations and Definitions of key terms	7
1.12 Organisation of the Dissertation	8
2 Literature Review	10
2.1 Chapter outline	10
2.2 Crime	10

2.2.1	Classification of crime	11
2.2.1.1	Visible crime	11
2.2.1.2	Occupational crime	11
2.2.1.3	Organised crime	11
2.2.1.4	Cybercrime	12
2.2.1.5	Victimless crimes	12
2.2.1.6	Political crime	12
2.2.2	Causes of crime	13
2.2.2.1	Gender and crime	14
2.2.2.2	Age and crime	15
2.2.2.3	Race and crime	16
2.2.2.4	Crime and education	18
2.2.2.5	Marital status and crime	19
2.2.3	Effects of crime	19
2.2.4	Crime prevention	20
2.2.5	Trends in crime	22
2.3	Poverty	22
2.3.1	Definition of poverty and inequality	23
2.3.2	Types of Poverty	24
2.3.2.1	Situational poverty	24
2.3.2.2	Generational poverty	24
2.3.2.3	Absolute poverty	24
2.3.2.4	Relative poverty	24
2.3.2.5	Urban poverty	25
2.3.2.6	Rural poverty	25
2.3.3	Causes of poverty	25
2.3.4	Measurement of poverty and inequality	26
2.3.4.1	Measurement of poverty	26
2.3.4.2	Measurement of inequality	28
2.3.5	Underlying issues of poverty variables	30
2.3.6	Poverty, unemployment and crime	31
2.4	Canonical Correlation Analysis	32

2.4.1	Matrix algebra and canonical correlation analysis	33
2.4.1.1	Covariance Matrix	33
2.4.1.2	Derivation of canonical variates	35
2.4.2	Objectives of Canonical Correlation Analysis	42
2.4.3	Designing a Canonical Correlation Analysis	42
2.4.4	Assumptions and impact on the analysis	43
2.4.5	Deriving the canonical functions and assessing overall fit . . .	44
2.4.6	Deriving canonical variates	44
2.4.7	Interpreting canonical functions	45
2.4.7.1	Level of significance	46
2.4.7.2	Magnitude of the Canonical Relationships	46
2.4.7.3	Redundancy and Canonical Correlation Analysis . .	46
2.4.8	Interpretation of the Canonical Variate	48
2.4.8.1	Canonical Weights	48
2.4.8.2	Canonical loadings	49
2.4.8.3	Canonical cross-loadings	49
2.4.9	Methods for interpreting the canonical variate	49
2.4.10	Limitations of the canonical correlation procedure	50
2.5	Concluding Remarks	51
3	Research Methodology	52
3.1	Chapter outline	52
3.2	Research Methodology and Design	53
3.2.1	Quantitative research	53
3.2.1.1	Rationale of the research design	54
3.3	Variables used in the research	55
3.3.1	Independent variables	55
3.3.2	Dependent variables	56
3.4	Research Questions and Hypothesis	57
3.4.1	Research questions	57
3.4.2	Research hypothesis	58
3.5	Area of the Study	58

3.6	Pilot Study	59
3.7	Target Population	60
3.8	Sampling Design	61
3.8.1	Sampling methodology	61
3.8.2	Sample size	63
3.8.3	Sample size calculation	64
3.9	Data Collection	68
3.10	Data Coding	69
3.11	Reliability and Validity	70
3.11.1	Reliability	71
3.11.2	Validity	71
3.12	Data Analysis and Interpretation	72
3.13	Research Ethics	73
3.14	Conclusion	74
4	Results and Discussions	75
4.1	Introduction	75
4.2	Data analysis and Interpretation	76
4.2.1	Overview of aggravated robbery crime	76
4.2.2	Interpretation of Canonical Correlation Analysis results	81
4.2.2.1	Testing the research hypothesis	82
4.2.2.2	Deriving canonical functions	82
4.2.2.3	Aggravated robbery crime canonical loadings analysis	83
4.2.2.4	Poverty canonical loadings analysis	85
4.2.2.5	Canonical cross-loadings of poverty-aggravated robbery crime datasets	87
4.2.2.6	Weights of the canonical variates	87
4.2.2.7	Redundancy analysis	89
4.3	Conclusion	91
5	Conclusions and Recommendations	97
5.1	Introduction	97

5.2	Conclusions	97
5.2.1	Empirical Findings	98
5.2.2	Theoretical implications	99
5.2.3	Policy implications	100
5.3	Recommendations	101
5.4	Concluding Remarks	102
	Appendices	113
	A Supervisor’s letter	113
	B SAPS Provincial and National letters	115
	C Data coding	118

List of Tables

1.1.1 Percentage distribution of population by population group	3
2.2.1 Comparison of convicted offenders by race	17
2.2.2 Aggravated statistics from 1994 - 2002	22
3.8.1 Limpopo aggravated robbery crime statistics: 01/01/2011 - 30/06/2013	65
4.2.1 Aggravated offenders per district	76
4.2.2 Statistics of poverty levels of offenders	79
4.2.3 Types of aggravated crime committed by offenders	80
4.2.4 Cluster comparison of aggravated robbery offenders	81
4.2.5 Multivariate Tests of Significance	82
4.2.6 Dimension Reduction Analysis	83
4.2.7 Eigenvalues and Canonical Correlations	84
4.2.8 Canonical Loadings for aggravated robbery crime set	85
4.2.9 Canonical Loadings for the poverty set	86
4.2.10 Cross-loadings for the poverty set	87
4.2.11 Cross-loadings for aggravated robbery crime set	88
4.2.12 Standardized canonical coefficients for the poverty set	88
4.2.13 Standardized canonical coefficients for the aggravated robbery crime set	89
4.2.14 Proportion of Variance of poverty set explained by its own Can. Var.	89
4.2.15 Proportion of Variance of aggravated robbery crime set explained by its own Can. Var.	90
4.2.16 Proportion of Variance of poverty set explained by aggravated robbery crime Can.Var.	91

4.2.17 Proportion of Variance of aggravated robbery crime explained by poverty Can. Var.	91
C.0.1 Coding format for poverty variables	119
C.0.2 Districts and clusters coding	119

List of Figures

2.3.1 Lorenz curve	29
2.4.1 Form of Data for Canonical Analysis	34
3.5.1 Limpopo Province Districts Map	60
3.8.1 Comparison of Sample size and Population	64
4.2.1 Comparison of types of aggravated crime per district	93
4.2.2 Aggravated crimes per cluster	94
4.2.3 Graphical comparison of types of aggravated crime	95
4.2.4 Canonical Correlation Analysis of Poverty and Aggravated robbery crime	96

Chapter 1

Background and Justifications

1.1 Introduction

Violent crime statistics, of which aggravated robbery is one, has aroused public interest and concern in South Africa. South Africa is among the countries with the highest levels of crime such as aggravated robbery crime(The Centre for Study of Violence and Reconciliation[CSVr], 2010). This has put the country on the spotlight. Empirical studies have been and are still being carried out with the objective of finding solutions of reducing crime. This research focused on aggravated robbery crime. Limpopo Province is not an exception since it also suffers from the bruises of this type of crime just like any other province of South Africa. It was this background that triggered this research to explore tentacles of aggravated crime related to poverty in Limpopo Province.

Robbery with aggravated circumstances has a great impact on economic development(for example, tourism such as at Kruger National Park, businesses and banks) and security of the residents of South Africa. According to Burger, Gould and Newham(2010), aggravated robbery has a major impact due to its violent nature. It has a direct impact in the sense that its effects are felt immediately. For example, stolen property and cash from business premises has instant effect. The indirect aspect is that it raises the cost of security, insurance and even limits business investment(Burger, Gould and Newham, 2010). Hence, goals of employment creation will

be far fetched.

The province where the research was conducted shares borders with Botswana, Zimbabwe and Mozambique. Limpopo Province also borders with provinces such as Gauteng, Mpumalanga and Northwest. This province has five districts, namely, Vhembe, Mopani, Capricorn, Waterberg and Sekhukhune. The province has a land area of approximately 123 910 km^2 and it comprises 10.2% of South Africa's land area according to Kyei and Gyekye's(2011) study. Table 1 shows the distribution of population per district as well as for the province by population group. Overall percentage of blacks from Census 2011 was 96.8% and other subgroups accounted for 3.2% of the provincial population(Statistics South Africa[StatsSA], 2012). Table 1.1.1 also exhibits comparisons of population groups in each district in relation to other districts for the selected previous censuses.

This province is marred by high poverty rates, unemployment rates and unequal distribution of income between population subgroups as indicated in the Municipal Report of 2012(StatsSA, 2012). From the report, unemployment was pegged at 39.5%, which was a decrease from the 1996 census where unemployment was 45.1%. Also, the dependency ratio was reported to be 67.3%, which measured the burden shouldered by a working population from the age group of 15–64 years. The stress of the responsibility of caring for others might be a driving factor for the employed to commit aggravated robbery. Illiteracy was 17.3% and it was one of the variables(educational attainment) which were investigated during the research.

The research analysed aggravated robbery crime by using a statistical methodology called Canonical Correlation Analysis(CCA). An in-depth study of poverty levels and aggravated crime would be of great benefit to stakeholders such as policy planners, the general South African public and in academic circles. This crime is the second-largest generator of other contact crimes and might be a result of circumstances such as social inequality, poverty, unemployment, gender perceptions, educational attain-

Table 1.1.1: Percentage distribution of population by population group

YEAR	District	Black African	Coloured	Indian/Asian	White
1996	Mopani	97.5	0.1	0.0	2.7
2001		97.5	0.1	0.1	2.3
2011		97.1	0.2	0.3	2.5
1996	Vhembe	98.5	0.1	0.2	1.4
2001		98.5	0.1	0.2	1.1
2011		98.3	0.1	0.4	1.1
1996	Capricorn	96.3	0.3	0.2	2.5
2001		96.3	0.4	0.3	3.0
2011		96.3	0.5	0.4	2.8
1996	Waterberg	90.9	0.3	0.1	9.6
2001		90.9	0.3	0.2	8.6
2011		91.5	0.5	0.4	7.6
1996	Sekhukhume	99.1	0.1	0.0	1.0
2001		99.1	0.1	0.1	0.8
2011		98.7	0.1	0.2	1.0
1996	Limpopo	97.0	0.2	0.1	2.8
2001		97.0	0.2	0.2	2.7
2011		96.8	0.3	0.3	2.6

Compiled from Census 2011 Municipal report(StatsSA, 2012)

ment, social exclusion and marginalisation, and so on(CSVR, 2010; South African Police Service Department of Police[SAPS], 2010). Hence, all stakeholders will find the analysis and conclusions useful.

1.2 The Research Problem

The overriding problem of rising of trio crimes(house robbery, street robbery and business robbery) in the society of Limpopo Province from 2009 to 2010 was of great concern to the public and the state(Burger and Gould, 2010). From statistical ev-

idence, as cited by Burger and Gould(2010), there had been an increase in street robberies(2%), house robbery(13.6%) and business robbery(19.1%). Hence, there was a need to explore the gap in knowledge of the relationship of this rising aggravated robbery crime with poverty. The objective of revealing some insights between poverty and aggravated robbery by using canonical correlation analysis was to come up with informed interventions to curb this crime. Another gap in knowledge was the inference of criminal offender records(docket analysis) by using canonical correlation analysis .

1.3 Purpose of the Study

The primary concern of this quantitative correlational study was to determine the relationship between poverty and aggravated robbery. The primary objective was to establish the significance of the relationship between these two variable sets in order to provide a well informed decision to policy planners and implementers of intervention strategies of aggravated robbery crime. Canonical correlation analysis revealed profound results in this correlational study. Research variables for poverty were gender, employment status, marital status, race, age and educational attainment. Aggravated robbery was classified as the dependent set with variables house robbery, bank robbery, street/common robbery, carjacking, truck hijacking, cash-in-transit and business robbery. These variables were analysed on a sample of 559 offenders who were sampled from Limpopo Province aggravated robbery offenders from January 2011 up to 30 June 2013.

1.4 Significance of the Study

The significance of the research lies on a new methodology of having long-term prevention measures of aggravated robbery crime based on the empirical findings of the research. When crime statistics are announced, nothing or less is mentioned about the effects of marital status on this type of crime and possible strategies of prevention along this ideology. In academic circles, this will fill the gap in knowledge about the practicability of canonical correlation analysis in analysing dockets with the recorded

criminal biographical information considered as poverty variables.

1.5 Objectives of the Research

The following research objectives were identified: (i) To examine the relationship between poverty levels and aggravated robbery crime in Limpopo Province.

(ii) To determine whether the relationships were significant between poverty levels and levels of aggravated robbery crime.

(iii) To quantify the strength of the relationship of poverty and aggravated robbery crime.

(iv) To identify the main components of aggravated robbery crime and poverty that describes crime and poverty in the province.

1.6 Research Questions

Research questions were formulated with the objective of the research in mind: to find solutions to the research problem. The study sought to answer the following main research question:

Is there any meaningful relationship between poverty and aggravated robbery crime in Limpopo Province?

The sub questions which were used to answer the main research question above were:

(i) Is there any correlation between poverty and aggravated robbery crime criteria?

(ii) In a set of variables of aggravated robbery, which variable has the most and which one has the least impact on creating a meaningful relationship between poverty and aggravated robbery crime?

(iii) In a set of variables of poverty, which variable has the most and which one has the least impact on creating a meaningful relationship between poverty and aggravated robbery crime?

1.7 The Research Hypothesis

The research hypothesis was:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_s = 0$$

$$H_1: \rho_i \neq 0 \text{ at least one } i=1; 2; \dots; s$$

and $s = \min(p, q)$ where p and q are variables of the two datasets.

This implies that all canonical correlations between linear combinations of the predictor set and the criterion set are zero. Statistical tests were performed to test the hypothesis after the analysis was run on a software package known as IBM SPSS (Statistical Package for Social Sciences) Statistics Version 22(2013). This enabled the interpretation of correlations between the poverty-aggravated robbery data.

1.8 Overview of Methodology

The research sought to answer the research questions by using quantitative correlational methodology. The research was done on the sampled districts, that is, Mopani, Vhembe and Capricorn of Limpopo Province in South Africa. The multistage sampling procedure which was used in the analysis left out Waterberg and Sekhukhune districts. Sample size was determined by using Krejcie and Morgan's (1970) formula. This was to safeguard the reliability of results since small samples has the likelihood of undermining reliability of results. The sample was drawn from a list of aggravated robbery offenders from 2011 up to June 2013 using SAPS software known as Business Intelligence System (BIS). Lastly, the sampled data was tested in IBM SPSS Version 22 program by using Canonical Correlation Analysis.

1.9 Assumptions of the Study

The study was based on the following assumptions:

- (i) The sample should be a true representative of the population about which it intended to make some generalisations. This assumption was justified by performing a multistage sampling design during the selection of offenders for data analysis.
- (ii) The data should be normal. This assumption of normality was achieved by hav-

ing a large sample based on Krejcie and Morgan's(1970) principles on sample sizes. Statistical methods such as Canonical Correlation Analysis are sensitive to sample size. Hence, the sample size should be sufficiently large for the success of Canonical Correlation Analysis methodology in determining interrelationships between the predictor set and criterion set.

(iii) The data obtained from SAPS data base of convicted offenders was assumed to be correct and reliably documented. SAPS official statistics were assumed to be correct and they were selected randomly by simple random sampling.

1.10 Limitations and Delimitations

Factors(limitations) which were not controlled include the following:

- Unavailability of information in dockets compromised the results. A remedy to this limitation was that such cases were excluded during data collection.
- Due to small samples available for the study, results were difficult to generalise. To overcome the problem, the sample per cluster was calculated using a method of calculating sample size which was proposed by Krejcie and Morgan(1970).

The study was delimited to Limpopo Province and crime on aggravated robbery. Results were limited to the models available in the SPSS software package. These enabled generalisability of results.

1.11 Abbreviations and Definitions of key terms

This section provides abbreviations and definitions of terms used in this research. The abbreviations and terms include the following:

AGE - Age

Aggravated robbery: It means unlawfully taking someone's property by inflicting bodily harm or injury.

BAROB - Bank robbery

BIS refers to Business Intelligence System.

BROB - Business robbery

CAS refers to Crime Administration Systems.

CARJ - Carjacking

CCA - Canonical Correlation Analysis

CID refers to Criminal Investigation Department.

CIT - Cash-in-transit

Contact crime: It refers to when there is physical contact between the perpetrator and the victim. For example, when a perpetrator hijacks a car, rapes the victim repeatedly and kills the victim afterwards (Crime Report 2010/2011)

CROB - Street/common robbery

EDUA - Educational attainment

EMPS - Employment status

GENDER - Gender

HROB - House robbery

MSTAT - Marital status

n.d: It is an abbreviation for “no date” for cited sources without dates.

Offender: The word refers to a person who transgresses moral or civil law (Audio-English.net Online Dictionary., n.d).

Quantitative research: In this research it refers to a research based on information derived from administrative records.

RACE - Race

SAPS refers to South African Police Service.

Street/public robbery or common robbery refers to robberies that take place on the streets and in other public or open spaces.

TRUH - Truck hijacking

1.12 Organisation of the Dissertation

Besides Chapter 1, this dissertation also focused on literature review (Chapter 2) where the underlying theories of crime were discussed. Chapter 2 also discussed the multivariate technique used in this research called Canonical Correlation Analysis.

After laying out the framework of the study, Chapter 3 explained the methodology used to uncover the relationships between poverty and aggravated robbery crime. Furthermore, it provided an explanation for the rationale of the approach, data collection and analysis methods. Sampling techniques such as multistage sampling, population and the geographic setting of the research are some of the items which were discussed in depth. Results and discussions are explained in Chapter 4. The analysis was run on SPSS and the chapter contains the outputs of the analysis and interpretations. Lastly, Chapter 5 dealt with conclusions and recommendations. Gaps in knowledge were filled but giving provisions of taking this research further again because not all gaps were filled.

Chapter 2

Literature Review

2.1 Chapter outline

This chapter provides an extensive review of literature of the research related to crime and poverty. The extensive literature provides a firm base for the research. It also focuses on statistical techniques which were used to determine the relationship between poverty and aggravated robbery in Limpopo Province. It discusses theoretical aspects of crime, poverty and Canonical Correlation Analysis. It is an area which has been researched as it is evidently supported by a rich literature. However, this research had another dimension since it sought to establish the relationship between crime and poverty in Limpopo Province in South Africa with an aim of shading light to policy-makers and the general public.

2.2 Crime

Crime is a prominent issue in any society since it retard economic growth of a country as well as threatening the security in a society. Hence, governments has the jurisdiction in defining it with the main aim of apprehending the perpetrators. From the criminologist's point of view, crime is a violation of criminal laws of a state(Schmalleger, 2009). Hence, crime is unique since it depends on the constitution of a country. It is reaffirmed by Tshiwula(1998) who noted lack of uniformity and crime dependence on the social differential environment.

However, a full understanding of the causes and prevention measures of crime, relies on the theories of crime in general. Crime is categorised into types and the list is not exhaustive since new types of crime crop up as society changes (Cole, Smith and Jong, 2012). Hence, crime can be classified as visible crime, occupational, organised crimes, and so on. The next section, used this methodology in classifying crime.

2.2.1 Classification of crime

Cole, Smith and Jong (2012) noted the following six types of crime. This is done in accordance with the nature of the act and the classification is not exhaustive.

2.2.1.1 Visible crime

This type of crime is visible and it includes crimes ranging from shoplifting to homicides. Furthermore, visible crime is classified into three types namely, violent crime, property crimes and public order. Violent crime constitutes crimes that cause bodily harm or death. This category includes crimes such as criminal homicide, forcible rape, aggravated robbery (Schmallegger, 2009). The second type of violent crime is property crimes which includes shoplifting, embezzlement and burglary. This crime normally causes its victims to suffer from emotional distress. The last category is called public order crime which constitutes crimes such as public drunkenness, vandalism and disorderly conduct (Cole, Smith and Jong, 2012).

2.2.1.2 Occupational crime

Occupational crime involves illegal acts which are done in the context of legal business or profession (Cole, Smith and Jong, 2012). An individual commits criminal acts by using his/her position. Examples of this crime can be cited as selling products of poor quality to consumers, a police officer removing court evidence by exercising government powers, theft by an employee, and so on.

2.2.1.3 Organised crime

Organised crime is unlawful activities that are committed by an organised syndicate (Schmallegger 2009). Such groups have rules and structures. Bank and cash-in-

transit robberies which are types of aggravated robbery crime belongs to organised crime. In terms of intelligence, the crime is highly organised (Schmallegger, 2009).

This type of crime is highly prevalent in South Africa and it ranges from criminal activities such as drug trafficking, human trafficking, firearms, bank robberies, smuggling of precious items, stolen vehicles, poaching, and so on (ODC Country Profile, 2002). However, countermeasures such as drafting new laws, improving policing strategies, asset forfeiture operations, and investigative and prosecutorial structures are being implemented (ODC Country Profile, 2002).

2.2.1.4 Cybercrime

Cybercrime is also one of the types of crime. It is a criminal offence to commit acts which violate laws by using computers and the Internet (Cole, Smith and Jong, 2012). They pointed out that these acts include violation of morality, theft by stealing information or funds, releasing Internet viruses, and so on. Presently this type of crime has made some inroads in South African society.

2.2.1.5 Victimless crimes

Victimless crime involves offenses whereby the victim and an offender privately exchange illegal goods or services willingly (Cole, Smith and Jong, 2012). Both the victim and an offender will be having an objective of benefiting. Examples of this crime are gambling, prostitution and illegal drug sales.

2.2.1.6 Political crime

According to Hagan (2011), political crime refers to offenses committed for ideological purposes. There are two types of political crime, that is, crime against government and crime by government. Hagan (2011) argued that crime by government includes violation by secret police, abuses of human rights and genocide. Ross also shared the same view that even governments can use state machinery to turn against its people and this is called genocide (Ross, 2012). An example of crime against government which can be cited is US 9/11 (11 September 2001) terrorism and 26 February 1993

incident where al-Qaeda members placed a bomb in the underground parking garage of the World Trade Centre in New York City(Ross, 2012).

2.2.2 Causes of crime

From Weatherburn's(2001) point of view, the subject of causes of crime is viewed as a complex topic. Criminologists have different views of factors that causes crime. Even though criminologists have divergent views on causes of crime, the need of crime reduction depends on sound knowledge of causes of crime(Bernard, Snipes and Gerould, 2010). Hence, scientific theories in criminology such as causal theories gave explanations of the causation of crime by highlighting the existence of causal relationships between variables such as delinquency and discipline (Bernard, Snipes and Gerould, 2010).

However, there is no consensus on the merit of crime causation theories although there is growing evidence that these theories can explain the risk of an individual to criminal offending(Strategic Policy Brief, 2009). Theoretical crime perspectives are embedded into six theories, that is, biological, developmental life-course, psychological, sociological, geographic and economic theories. Biological theories focus on inherited genes, evolutionary factors and brain structure. The article proposed prevention measures based on this theory such as maternal health initiatives and reducing alcohol consumption since this causes neurological damage and lack of long-term cognitive skills(Strategic Policy Brief, 2009).

Geographic theories focus on geographic distribution of crime and propose modification of the environment as a means of preventing crime(Strategic Policy Brief, 2009). For example, increasing police patrols in high density areas and by making physical improvements in areas which are prone to crime. However, Weatherburn (2001) pointed out that causation is not significant when dealing with crime since what criminologists regard as causing criminal behaviour does not result in it. The so-called factors of crime increase the risk of criminal behaviour. So, the more risk factors an

individual has, the greater the probability of being involved in crime.

This brings about the discussion of factors which may increase criminal behaviour according to the above assertion. The same conception is shared by Schmalleger when a criminal event is a result of an offender with the following attributes: personal life experiences, peculiar biological traits, a distinct personality, personal values and beliefs(Schmalleger, 2009). The following are the few basic factors of crime. The list of factors is not exhaustive.

2.2.2.1 Gender and crime

Gender is regarded as the strongest best single predictor of crime(Schmalleger, 2009). In addition, crime is regarded as having a correlation with gender(Bernard, Snipes and Gerould, 2010). Siegel(2010) pointed out that the gender gap(differences of male and female crime rates) can be attributed to emotional, physical, psychological differences between males and females. According to the sociological theory of crime, gender gap can be viewed as caused by cultural factors such as early socialisation and role expectations(Schmalleger, 2009). Hence, this explains the gender gap disparity.

Bernard, Snipes and Gerould(2010) cited feminist criminologists Adler and Simon who argued that there would be an increase of female offenders if women moved out from traditional societal roles. They argued that Adler's theory proposed that female offenders would commit violent crimes whereas Simon advocates that the crime would be predominantly white-collar crime (Bernard, Snipes and Gerould, 2010). They suggested that Simon's theory might be correct (Bernard, Snipes and Gerould, 2010). Besides the nature of crimes predominantly committed by females, the discussion cannot be closed without an in-depth discussion on reasons for low crime rate of women against that of men.

The power-control theory outlines that parents control girls more than boys and hence this might result in boys engaging in criminal activities more than girls(Bernard,

2010). It is more significant in families where the father has more power than the mother because of his economic status. Now, considering families where the father and mother have equal powers, the rate of criminal activities will be least. An alternative explanation has been given by using the strain-type theories(Bernard, Snipes and Gerould, 2010).

There are various theories that gave explanations to why men's crime rate is higher than women's crime rate when comparing the two criminal rates. From a biological perspective theory, some traditional criminology theories attribute this difference to unequal hormone levels such as testosterone which is higher in males than in females(Bernard, Snipes and Gerould, 2010). Summing this discussion of gender and crime, more researches are coming up with more explanations to gender and crime perspective.

2.2.2.2 Age and crime

From the United States, crime statistics showed that youths who were between thirteen and eighteen years made up about 6% of the population and approximately 30% were arrested for major crimes(Schmallegger, 2009). These youths constituted about 18% of the total arrests for the whole population in a given year (Schmallegger, 2009). These statistics highlighted relationship of age and crime. This brings up Siegel's(2010) point that age is inversely related to criminality and it may be based on biological characteristics(from the biological theory's perspective). The author argued that young people were still strong than adults and had an advantage when it came to vigorous violent crimes due to high levels of testosterone they had as compared to adults(Siegel, 2010). This testosterone is linked to aggression and violence(Siegel, 2010). Hormones level declines as one grows up and this explains why violence diminishes with age(Siegel, 2010). According to Schmallegger's(2009) point of view, the hormonal level start to decrease at an age of 25 years. This suggests a relationship between age and crime.

2.2.2.3 Race and crime

Race cannot be left out as a contributing factor to criminal activities. There are a number of theories based on empirical findings that explain the correlation between race and crime. For example, theories which highlight this subject are sociological, biological and cultural conflict theories(Greene and Gabbidon, 2012). Biological theory of race and crime coined the causes of crime on the basis of skin colour. In Gabbidon's(2010) paper, race was defined as classification of people on the basis of their skin colour and cultural characteristics. From a biological perspective, Gabbidon(2010) cited a research of army personnel from the southern portions of Italy which were considered inferior in terms of laziness and inclination to criminality. This view was based on the fact that southern Italy was mostly inhabited by Africans. Africans were viewed as having heads which were undeveloped at the back and weighed less compared to those of Whites(Gabbidon, 2010). Physical characteristics were linked to criminal behaviour(Bernard, Snipes and Gerould, 2010). Based on these findings, Gabbidon(2010) argued that most of the homicides in Calabria, Sicily and Sardinia were mostly due to African and Oriental elements. Schmallegger shared the same view as it is depicted by the inclusion of crime statistics of the United States of 2006(see Table 2.2.1)(Schmallegger, 2009).

The conviction of offenders by race(Table 2.2.1) explained the great disparity between Blacks and Whites. According to Gabbidon(2010), it showed an existence of correlation between race and crime. From empirical findings, it was deduced that Black Americans had high criminal rates compared to Whites(Gabbidon, 2010). It has been argued that Black children attended disadvantaged schools; and as a result, there will be less chances of correction of the learning disabilities(Gabbidon, 2010). Research had shown that the disparity in educational resources contributes much to underemployment and criminal behaviour(Gabbidon, 2010).

From the biological perspective on crime, it is alluded to that Black males are more muscular than their White male counterparts(Greene and Gabbidon, 2012). Accord-

Table 2.2.1: Comparison of convicted offenders by race

Arrest Rates by race (PER 100 000 people) and Relative Rates of Arrest for Selected offenses, 2006			
Offenses	African Americans	White	Rate of Arrest, African Americans vs. White
Murder	13	2	6.5:1
Rape	19	6	3.2:1
Robbery	119	16	7.4:1
Aggravated assault	321	100	3.2:1
Burglary	163	65	2.5:1
Larceny	725	251	2.9:1
Motor vehicle theft	118	28	4.2:1
Weapons	123	33	3.7:1
Drug abuse	1 086	331	3.3:1
Prostitution	68	16	4.2:1

Adapted from Schmalleger(2009)

ing to research findings, Black males had lower IQ (Intelligence Quotient) than White males and by this argument Black crime rates were expected to be higher than for Whites (Greene and Gabbidon, 2012). However, from a sociological perspective, crime was not viewed as being correlated with race. Greene and Gabbidon(2012) pointed out that from a sociological perspective, age, unemployment and poverty contribute much to crime. Furthermore, they attributed high rate of American Black offenders to racial discrimination(Greene and Gabbidon, 2012). The association of race and crime was also supported by contemporary social disorganisation theorists who discovered an association between unemployment of Black males and economic deprivation and violent crimes(Greene and Gabbidon, 2012).

Another explanation to the relationship of race and crime was through the application of the General Strain Theory(GST). From Greene and Gabbidon(2012), it is argued

that strain is caused by an injection of stimuli with a negativity characteristic in the environment. According to the same source, criminal behaviour is caused by negative stimuli such as child abuse and neglect, verbal threats, criminal victimisation and stressful life events (Greene and Gabbidon, 2012). In the South African perspective, authorities like Shaw (2002) attributed the disparity of race and crime to the apartheid regime. It is argued that the apartheid-imposed geographies and principles ensured that crime was concentrated in black residential areas (Shaw, 2002). By the same argument, the majority of the victims of violent crime and even offenders remain blacks. Limpopo Province shares the same fate.

2.2.2.4 Crime and education

Lochner and Moretti(2003) assert that education is regarded as one of the factors that decrease the probability of imprisonment. On the other hand, lower educational attainment levels will propel chances of incarceration(that is, imprisonment)(Alliance for Excellent Education[AEE], 2006). Alongside this perception, Lochner and Moretti(2003) also deduced from their studies that the differences in imprisonment gap rates between black and white men could be attributed to educational attainment(Lochner and Moretti, 2003). They found out those crimes such as murder, aggravated assault and vehicle theft were associated with education(Lochner and Moretti, 2003).

The correlation of crime and education was also shared by AEE(2006). Based on the theories which hypothesised that educational attainment reduces the rate of crime, the following reasons were noted from AEE(2006):

- (i) High wages of people with high educational qualifications will reduce the perception of committing crime.
- (ii) People with high educational attainment have a fear of being imprisoned.
- (iii) Classroom interactions instil societal values which oppose criminal behaviour.
- (iv) By keeping youths in class, this will reduce their ego to engage themselves in criminal activities since there will be no time for engaging in activities that violates societal norms.

2.2.2.5 Marital status and crime

Many scholars have researched on the correlation of marital status and crime. Studies on the impact of marital status showed that marriage reduced the probability of reoffending(Weisburd and Waring, 2001). This notion is also shared by Bernard, Snipes and Gerould, 2010) who point out that a reduction of crime is due to marriage. They argued that good, functional marriages will influence youths to desist from crime since the marriages instil moral values of the society(Bernard, Snipes and Gerould, 2010). Therefore, marriage was viewed as a role player in the reduction of the probability of committing crime.

2.2.3 Effects of crime

Crime everywhere has disastrous effects socially, economically and even on development. South Africa is not an exception and many people do not have physical safety and are living in trauma and fear(Kok, 1998). It prohibits business activities and business confidence by scaring away investors(Kok, 1998). In addition, crime diverts resources earmarked for economic growth, for protection purpose, exerts pressure on health facilities and creation of a non-conducive environment to productive activities(Bhorat and Kanbur, 2006). Hence, a high crime environment will breed low investment opportunities. By so doing, this will result in high unemployment rates(Kok, 1998). Another effect of crime is the migration of skills. Widespread emigration of professionals in the post-apartheid era in South Africa can be attributed to high crime rates which compelled them to leave the country for crime-free environments(Bhorat and Kanbur, 2006). Official statistics of 1994 showed that South Africa lost skilled labour in high-level occupations due to high levels of crime(Bhorat and Kanbur, 2006). Crime can also reduce the inflow of tourists due to negative publicity of crime in any particular country.

Many researches on impacts or effects of crime have focused on society, victims of crime, economy and many more. In line with this view, Detotto and Otranto (2010) says:

On the other hand, crime imposes great costs to the public and private actors, such as stolen and damaged goods, lost lives, security spending, pain and suffering. ...

From empirical findings, crime reduces the Gross Domestic Product(GDP) indirectly(Detotto and Otranto, 2010). Many governments will spend a lot of money to activities related to crime. Insurance companies also shoulder the burden of lost property costs. Crime also causes overload of the health system due overflow of victims of crime.

2.2.4 Crime prevention

It is an obligation of any government to institute some crime prevention programmes with an objective of reducing crime. The solution of crime prevention is wholly embedded in the root causes of crime(Emmett and Butchart, 2000). One of the methods which was proposed was to focus on early development of children aged from three to five years by home visiting, pre-school training and parent effectiveness training, that is, application of sociological perspectives of criminology(Emmett and Butchart, 2000).

School-based youth educational interventions were suggested to be of paramount importance for curbing crime(Emmett and Butchart, 2000). From the same source, it was suggested to use teachers as intervention agents in order to solve problems of anger management, conflict resolutions and life skills training. The Washington Community Violence Prevention Programme was cited as an example(Emmett and Butchart, 2000). Another method will be to apply regulation intervention whereby dangerous weapons are reduced by the introduction of enforcement laws which restrict possession of guns without licences and carrying them in public(Emmett and Butchart, 2000). Emmett and Butchart(2000) also proposed another prevention method of popular engagement and empowerment by allowing people to give their input in dealing with the safety of their environment, allowing them to prioritise actions and what resources

should be allocated to interventions, imparting knowledge about injury control and to make their communities crime free.

Businesses need to be guarded in order to prevent criminal activities(Bressler, n.d). Bressler(n.d) outlined that business crime should be prevented by surveillance techniques such as employees, alarms or video monitors, locks or key control, computer firewalls, security lighting, employment policies, and so on. South Africa like many other countries, is enforcing some of the crime prevention programmes in order to combat crime. The 1996 National Crime Prevention Strategy(NCPS) reinforces the following:

- Improving the criminal justice system in terms of efficiency, enforcing laws and protecting human rights.
- Putting constraints on the environmental or situational opportunities for crime and by upgrading existing infrastructure.
- Encouraging community participation and totally involvement in crime prevention.
- Improving border controls, cross border crime and regional co-operation(Rauch, n.d).

The South African government improved the NCPS in order to have more grip in reducing crime. The 1998 White Paper on Safety and Security was approved by the South African cabinet and it focused mainly on law enforcement, crime prevention and institutional reform by clarifying accountability(Rauch, n.d).

The SAPS(2010) also suggested preventive measures for violent crimes like aggravated robbery and other crimes in general. The measures are as follows:

- Massive economic development and promoting growth in the informal sector in order to create employment and by so doing reducing the unemployment rate.
- Developing suitable human settlement where people can live a meaningful life which will result in the development of strong and positive self-esteem.
- Policing strategies geared to the reduction and elimination of crime, particularly

the trio crime rates of carjacking, home robbery and business robbery(SAPS, 2010).

2.2.5 Trends in crime

The South African crime statistics showed a downward trend of crime from 1994 to 1996(Burchell and Milton, 2005). There was a steady increase as from 1997 to 2001 and crime rates decreased in 2002. Thereafter, it went back to its 2001 level(Burchell and Milton, 2005). The trend of aggravated robbery which was being researched had been on the spotlight in South Africa because of its direct effects to business and society. It had shown a significant increase since 1994 and the Western Cape appears to have had the highest levels of aggravated robbery crimes(United Nations Office on Drugs and Crime[ODC], 2002 and Burchell and Milton, 2005). Usually assailants are armed and this upward trend was an indication of the availability of firearms and other weapons which were used on victims(ODC, 2002). Table 2.2.2 below shows the aggravated crime statistics of South Africa.

Table 2.2.2: Aggravated statistics from 1994 - 2002

Robbery (aggravated) rates: 1994/95 - 2001/02 (per 100,000 population)							
1994/95	1995/96	1996/97	1997/98	1998/99	1999/00	2000/01	2001/02
219.5	195.5	164.0	177.2	219.9	229.2	257.7	258.5

Source: Adapted from ODC(2002)

2.3 Poverty

Poverty is one of the distinct problems experienced by South Africa besides problems such as crime. This is viewed as a multi-dimensional concept(Bhalla and Qiu(2006)). There are divergent views on the definition of this societal problematic issue. This implies that it has different meanings in different cultures and at different levels of economic development(Bhalla and Qiu, 2006). This section explores several concepts of poverty such as definition, classification and causes of poverty as well as measurement of inequality. Furthermore, a justification of the determinants of poverty in this research is provided in this section.

2.3.1 Definition of poverty and inequality

In the context of poverty literature, poverty and inequality are correlated but these two phenomena are different. The correlation is noticeable in developing countries where they pose serious threats. Firstly, poverty is discussed followed by inequality.

There is no single definition of poverty in the poverty literature. Hence, there is no unanimous definition of poverty. However, poverty can be defined as a deprivation in well-being (Haughton and Khandker, 2009; Phogole, 2010). From this definition, poverty is viewed as a lack of command over commodities due to insufficient of income or consumption to some defined threshold (Bhalla and Qiu, 2006; Haughton and Khandker, 2009). An individual is regarded as poor if for example, the income is below the threshold. The definition distinguish the rich from the poor. It can also be defined as the ability to obtain specific type of consumption goods such as education, health, food, shelter, and so on (Haughton and Khandker, 2009). From another view point, poverty is conceptualised as the capability to function in a society where lack of the key capabilities such as income, education, freedom of speech and sense of powerlessness, will label an individual as poor (Bhalla and Qiu, 2006; Haughton and Khandker, 2009).

On the other hand, inequality is defined as the variation of living conditions over the entire population (Haughton and Khandker, 2009). In contrast to poverty, inequality does not focus only to a portion of individuals who are below the poverty threshold but it also focuses groups in a population or among countries. It can also refer to an imbalance in the distribution of essential resources such as income among a specific group or population (Studies in Poverty and Inequality Institute [SPII], 2007). Inequality is of paramount importance because it a significant factor to poverty and crime. However, with these definitions in mind, poverty is classified in this research using the categories in the next section below.

2.3.2 Types of Poverty

2.3.2.1 Situational poverty

Situational poverty is defined as lack of resources which is generally caused by a sudden crisis or loss, that is, death of a breadwinner, environmental disasters, divorce, chronic illness, and so on(Jensen, 2009). This type of poverty is temporary and often people embedded in this type of poverty often have a negative attitude towards charity(Jensen, 2009).

2.3.2.2 Generational poverty

This type of poverty is different from situational poverty. It occurs in families which have been in poverty for at least two generations or have been born into poverty(Jensen, 2009). Researches have revealed that families trapped in this type of poverty have difficulties in getting out of it(Jension, 2009).

2.3.2.3 Absolute poverty

It is based on the minimum requirement needed to sustain life such as shelter, running water and food (Jensen, 2009; Govender *et al.*, n.d). This deprivation of minimum requirements for the maintenance of the physical body is usually measured by the “shopping basket” which is based on calculation of the income required to sustain the minimum calorie required by an individual(SPII, 2007). In addition, people in absolute poverty are mostly preoccupied with day-to-day survival(Jensen, 2009).

2.3.2.4 Relative poverty

Jensen(Jensen, 2009) defined relative poverty as the economic status of a family whose income is insufficient to meet a societal average standard of living. This brings in the concept of comparison which is used to determine the poverty level relative to the living standards of a society(Govender *et al.*, n.d). In other words, the relative poverty of a person is determined by comparing the standard of living with that of the society.

2.3.2.5 Urban poverty

There are many definitions of urban poverty. According to Masika, Haan and Baden(1997), there is no single definition of urban poverty. From Jensen's(2009) view point, it occurs in metropolitan areas with populations of at least 50 000 people. Urban poverty comprises acute stressors such as overcrowding(slums), violence, low life expectancy, nutrition, food budget and sanitation against the common index of material welfare(Masika, Haan and Baden, 1997; Jensen, 2009). The identification of urban poor can be done by using approaches to measure urban poverty(Baker and Schuler, n.d). Baker and Schuler(n.d) proposed a method known as money-metric measures which used income or consumption to assess the affordability of a basic basket of food by a household at a given time.

2.3.2.6 Rural poverty

Rural poverty occurs in non-urban areas with populations below 50 000 people(Jensen, 2009). The most common feature is that there are more single-guardian households and the creation of job opportunities is problematic(Jensen, 2009). The rural poverty rate is increasing and Jensen(2009) argues that it has exceeded the urban rate every year according to the US data. Measurement procedures are the same as for urban poverty.

2.3.3 Causes of poverty

Factors which cause poverty are intertwined. Growing inequalities deprive the society from a command on nutrition, medical expenses, decent shelter and participation in society(Southern African Regional Poverty Network[SARPN], n.d). According to Phogole(2010), poverty in South Africa is high. This might have been caused by the repressive apartheid laws which promoted a skew distribution of economic assets(SARPN, n.d). Apartheid policies gave Black South Africans inferior education and no rights of land ownership(SARPN, n.d). Examples of repressive laws as adapted from SARPN(n.d) are:

- The Land Act of 1913, which confined the land area that Africans could legally own or rent to 13% of South Africa.

- The 1913 Mines Act, which contained the first of many jobs reservations policies.
- The Urban Areas Act of 1945.
- Coloured Preferences Policy.
- The Group Areas Act of 1956, which restricted African access and African economic activity in the urban areas.

Generally, poverty may be caused by low agricultural productivity and output which relies heavily on over dependency on rainfall(Phogole, 2010). In South Africa, high unemployment and inequalities in incomes have been noted as causes of poverty(Phogole, 2010). For example, the Expanded Public Works Programme(EPWP) has not been able to create new jobs(Phogole, 2010). Also, lack of adequate basic services which might be propelled by budgetary constraints of governments might be a contributory factor to poverty(Phogole, 2010). Other factors which have been cited in the document are the rise in food prices and the high population growth which offset government plans to reduce the number of poor people(Phogole, 2010). Phogole(2010) also argued that internal conflict or civil war will result in failure of the rule of law which will drive out the basic economic incentives. For example, war in the Democratic Republic of Congo has resulted in the economic mismanagement which has made about 50 million people unstable(Phogole, 2010).

2.3.4 Measurement of poverty and inequality

2.3.4.1 Measurement of poverty

One of the goals of poverty measurement is to give direction to policy implementation. Methods of poverty measurement are established after the setting out of the minimum acceptable indicator of welfare(the poverty line) which can separate poor from those who are not poor(Haughton and Khandker, 2009). There are many methods of measurement of poverty in the literature of poverty. Some of the methods are the headcount index, poverty gap index, squared poverty gap index and the Watts index to mention just a few(World Bank Institute[WBI], 2005). Three methods are discussed below. This is due to space because the list is long. The selected methods

are discussed below using WBI(2005)'s presentation unless stated otherwise.

(a)Headcount index

This method is widely used and it is given by the following formula

$$P_0 = \frac{1}{N} \sum_{i=1}^N I(y_i < z) \quad (2.3.1)$$

where P_0 is the proportion of the poor in a given population, N is the population size, y_i is the income(or expenditure), z is the poverty line and $I(.)$ is an indicator function which takes a value of 1 if true or 0 otherwise.

Its popularity is because it is not difficult to construct and understand(WBI, 2005). Even though it is widely used, it does not mean that it is free from shortcomings. It violates the transfer principle, which can be stated as the transfer of income from a poorer person to a richer person should increase the headcount index(WBI, 2005).For example, a poor household were to give to a better off household, the headcount index will either be unchanged or it can decrease. It also fails to capture the intensity of poverty.

(b)Poverty gap index

Another popular measurement of poverty is known as “poverty gap index”. It is an average of shortfall of incomes of all individuals that fall below the poverty line. The poverty gap index is expressed as a percentage of the poverty line. It is denoted by P_1 . Firstly, the income gap is given as

$$G_i = (z - y_i).I(y_i < z) \quad (2.3.2)$$

Then the poverty gap index(P_1) formula may be written as

$$P_1 = \frac{1}{N} \sum_{i=1}^N \frac{G_i}{z} \quad (2.3.3)$$

For the non-poor population group, poverty gap will be zero.

(c)The Watts Index

This measurement was proposed by Watts(WBI, 2005). The formula is of the form

$$W = \frac{1}{N} \sum_{i=1}^q [\ln(z) - \ln(y_i)] \quad (2.3.4)$$

where N = indexed individuals in the population in ascending order of income(or expenditure),

q = individuals with income(or expenditure) y_i falls below z .

It is widely used by researchers and it satisfies the theoretical aspects in a poverty index(WBI, 2005).

2.3.4.2 Measurement of inequality

Inequality is viewed as having a significant effect on crime and poverty. This is a result of the variation of living conditions over the entire population mainly due to disparity in distribution, for instance, in income and expenditure. Inequality measures are often applied to many dimensions of living standards such as expenditure, malnutrition, income, education, land, assets, and many others where a continuous distribution is shown(WBI, 2005). The most popular inequality measurement is the Gini coefficient. However, there are other measurements of inequality in the literature of inequality such as Theil's T index, the Coefficient of variation, Atkinson's inequality measures and the McLooze index. The Gini coefficient and Theil's T index measures of inequality are described below.

(a)The Gini coefficient

Its construction is based on the Lorenz curve, cumulative percentage of the population and cumulative percentage of a specific resource, for instance, income(or consumption)(WBI, 2005). In addition, observations should be ordered. The horizontal axis represent the cumulative percentage of the population while the vertical axis represent the cumulative percentage of resource, for example, consumption. Then the Lorenz curve is a result of plots of ordered pairs of the cumulative percentage of the population and the cumulative percentage of a resource(for example, expenditure). Furthermore, a 45 degree line is drawn on the diagram . The general shape is like the one depicted on Figure 2.3.1(WBI, 2005). The diagonal line, that is, the 45 degree line, shows perfect equality(WBI, 2005). Then the Gini coefficient is obtained by dividing area between the 45 degree line and the Lorenz curve by the

area below the diagonal. From Figure 2.3.1, it is given as $\frac{A}{A+B}$ where A represent area between the diagonal line and the Lorenz curve, and B is area below the Lorenz curve(WBI, 2005). It takes a minimum value of zero(perfect equality) and maximum value of one(perfect inequality).

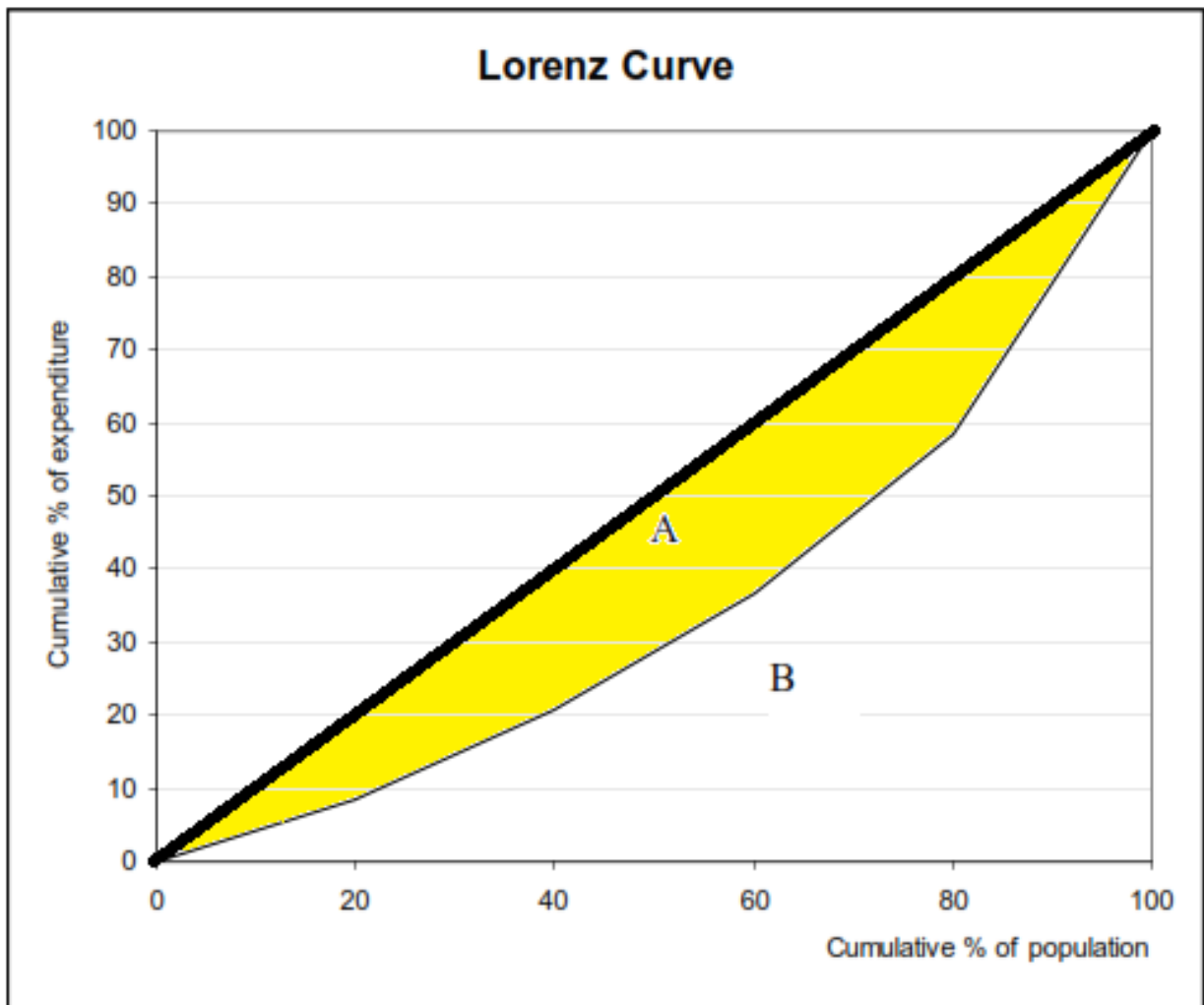


Figure 2.3.1: Lorenz curve

Adapted from WBI(2005)

The area below the Lorenz can be approximated by using trapezoids. Let x_i and y_i be points, on the horizontal and vertical axis respectively. Then from WBI(2005), the Gini coefficient of inequality formula is expressed as

$$G_1 = 1 - \sum_{k=1}^N (x_k - x_{k-1})(y_k + y_{k-1}) \quad (2.3.5)$$

The formula gives an approximation of G_1 and there are other formula which can improve the approximation. Statistical softwares(for example, Stata and SPSS) can be used in the computation of this Gini coefficient.

(b)Theil's T index

The Theil's T index belongs to the family of the Generalised entropy measures of inequality(WBI, 2005). The formula of the Generalised entropy measures is given as

$$GE(\alpha) = \frac{1}{N\alpha(\alpha - 1)} \sum_{i=1}^N \left[\left(\frac{y_i}{\bar{y}} \right)^\alpha - 1 \right] \quad (2.3.6)$$

for real values $\alpha \neq 0, 1$;

where \bar{y} is the mean income(or expenditure).

Therefore, the Theil's index formula is of the form(WBI, 2005)

$$GE_1 = \frac{1}{N} \sum_{i=1}^N \left[\frac{y_i}{\bar{y}} \cdot \ln \frac{y_i}{\bar{y}} \right] \quad (2.3.7)$$

Again this index can be computed by using statistical softwares.

2.3.5 Underlying issues of poverty variables

This study examined the relationship between selected poverty variables with those of aggravated robbery crime for jailed offenders of the crime. The complexity of the causes of poverty translates to the difficulties of coming up with poverty variables. However, Haughton and Khandker(2009) has identified possible determinants(levels or variables) of poverty such as demographic and socio-economic variables, for example, unemployment, gender of head of household, educational attainment, dependency ratio, and so on. These variables are correlated with poverty as shown by previous researches(Haughton and Khandker, 2009; Ebenezer and Iyaniwura, 2012).

Despite the complexity of poverty, it has been deduced from many researches that poverty is correlated with living conditions such as education, employment and health.

Hence, in poverty analysis, gender, age, marital status and race have sizable effects on the likelihood of poverty of an individual. It also implies that poverty can be measured without the use of population's income level and food security. Based on the above argument and in line with other researches, demographic and socio-economic variables were taken as poverty indicators. These were age of the offender, marital status, educational attainment, race, gender and employment status. The choice of the variables in this secondary analysis was in line with how data of offenders was recorded in dockets by the South African Police Services(SAPS). SAPS do not record income of offenders.

2.3.6 Poverty, unemployment and crime

The relationship among poverty, unemployment and crime has propelled more researches in this dimension. But there have been some disagreements on the issue. From some empirical studies, it has been noted that unemployment is a cause of poverty and it then causes crime(Bernad, Snipes and Gerould, 2010). Bernard, Snipes and Gerould(2010) used Glaser and Rice's results from their researches that crime was found to be inversely related, that is, crime is high when unemployment rate decreases on the basis that unemployed parents will be available all the time to take care of their children. From the crime and unemployment's perspective, they suggested that both variables are directly related, that is, if crime is high then unemployment will also be high(Bernard, Snipes and Gerould, 2010). Recent studies have shown that when unemployment is high, this will imply that crime rate is going to increase(Bernard, Snipes and Gerould, 2010). However, Bernard, Snipes and Gerould(2010) used results analysis of Chirico's empirical studies which showed that there was a positive correlation between unemployment and crime only for smaller samples which might be different for large samples.

Studies have also shown that economic inequality, that is, poverty is regarded as being associated with crime(Bernard, Snipes and Gerould, 2010). To support this argument, it was found that areas which were poverty-stricken and with high unemployment had less crime rates but criminal activities were committed to wealthy and employed

people by the poor and unemployed people(Bernard, Snipes and Gerould, 2010).

A thorough analysis of all aspects mentioned above about poverty and crime, were analysed by using one of the methods of multivariate techniques known as canonical correlation analysis. Deductions were made by using this rich literature of crime and poverty. This powerful method which was first coined by Hotelling in 1936 examined patterns of interrelationships between sets of variables of aggravated robbery and poverty(Levine, 1977). Aspects of canonical correlation analysis are discussed below.

2.4 Canonical Correlation Analysis

From the above brief outline, CCA is a multivariate technique which has a sole objective of identifying the nature of relationships within one set of variables and between two sets of variables. The two sets of variables are considered as a set of dependent variables and the set of independent variables. However, results of relationships are never compromised by interchanging the names of the two sets because CCA is symmetric. Also the relationships between two datasets are explained by a pair of linear combinations from these sets. In order to have legitimate results, the two sets of variables should measure the same sampling unit(Rencher, 2002). CCA is different from regression. Besides being used in modeling relationships, regression is used in prediction with a set of predictor variables and a response variable. Also, it can deal with many predictor variables. The method is called multiple regression. This deals with many-to-one relationships(many predictors are compared with one response variable) whereas CCA deals with many-to-many relationships, that is, many independent variables to a multiple of dependent variables.

Canonical correlation analysis generates canonical functions which are independent and maximises correlation between the two datasets. Each canonical function has two canonical variates; one from the predictor set and the other from the criterion set(Hair *et al.*, 2006). Then canonical functions are computed in such a way that the coefficient of correlation between the two datasets is maximised. This can be done

by using matrix algebra and statistical programs for large data sets. The first pair of canonical variates (first canonical function) account for the maximum of correlation between the two sets and the second pair to be derived will account for the residual variance, that is, the variance which was not accounted for by the first canonical function (Hair *et al.*, 2006). Hence, this process is done repeatedly and the first canonical function exhibits the highest canonical correlation. Canonical correlation coefficient measures the strength of relationship between the two canonical functions. Then the maximum number of canonical functions is equivalent to the number of variables in the smaller set which might be the set of independent variables or the set of dependent variables (Manly, 1986).

2.4.1 Matrix algebra and canonical correlation analysis

The mathematical background of canonical correlation is based on matrix algebra which involves aspects such as matrix inversion, orthogonal vectors, partitioned matrices, eigenvalues, eigenvectors and linear combinations (Manly, 1986). Manly (1986) and Rencher (2002) had a rich mathematical background of canonical correlation analysis.

Firstly, available data for analysis is represented on a data matrix. Figure 2.4.1 (Levine, 1977) shows a data matrix used for canonical correlation analysis. The sets \mathbf{X} and \mathbf{Y} are not necessarily a representation of independent and dependent variables. Each set is composed of variables such as, x_1, x_2, \dots, x_p , and y_1, y_2, \dots, y_q , where p is number of variables in set \mathbf{X} and q are variables in set \mathbf{Y} (Levine, 1977).

2.4.1.1 Covariance Matrix

Before expressing the general form of the sample covariance matrix there is need to state formulas for calculating the sample mean and sample variance of univariate variables, sample covariance and sample correlation of bivariate variables. Suppose there are n observations for a random variable x_i , where $i = 1, 2, \dots, n$. The sample mean is defined as

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.4.1)$$

Figure 2.4.1: Form of Data for Canonical Analysis

	$x_1 \ x_2 \ \dots \ x_p$	$y_1 \ y_2 \ \dots \ y_q$
case 1	X set	Y set
.		
.		
case N		

The sample mean is then used in the calculation of the sample variance. The sample variance of a univariate variable is given as

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (2.4.2)$$

where \bar{x} is obtained from equation (2.4.1). For the bivariate case where two random variables say x and y measure each research unit the bivariate sample variance(using (2.4.2)) is defined as

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (2.4.3)$$

where the sample means \bar{x} and \bar{y} are obtained from (2.4.1). Then equations (2.4.2) and (2.4.3) are used to find the sample correlation. The sample correlation is

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.4.4)$$

Individual entries of the covariance matrix for multivariate data is obtained by using equations (2.4.1) to (2.4.4). The sample covariance matrix or variance-covariance matrix is a matrix of sample covariances. Suppose there are t variables(from Section 2.4.1 , for example, it will be $t = p + q$) for a given multivariate data matrix. The

sample covariance matrix $\mathbf{S}=(s_{jk})$ is then given as

$$\mathbf{S} = \mathbf{s}_{jk} = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1t} \\ s_{21} & s_{22} & \cdots & s_{2t} \\ \vdots & \vdots & \ddots & \vdots \\ s_{t1} & s_{t2} & \cdots & s_{tt} \end{pmatrix} \quad (2.4.5)$$

From the above covariance matrix (equation (2.4.5)), the sample covariances for the j th variables are found on the diagonal. Then entries off the diagonal are for pairwise sample variances. The sample covariance of equation (2.4.5) of the j th variable, $s_{jj}=s_j^2$, is calculated as in equation (2.4.2). By using the j th column of the data matrix with t variables, the s_{jj} entry is defined as

$$s_{jj} = s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (2.4.6)$$

where \bar{x}_j is the mean of the j th variable. For each of the pairwise sample covariances of the j th and k th variables, the entry s_{jk} is calculated as in (2.4.3). Such type of entries are obtained by the following formula:

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad (2.4.7)$$

It should be noted that the subscript i corresponds to units measured by the variables in the sample under study. The variance-covariance matrix \mathbf{S} , is symmetrical since from (2.4.7) $s_{jk} = s_{kj}$. Thus, all entries are obtained by applying (2.4.6) and (2.4.7).

2.4.1.2 Derivation of canonical variates

The derivation of canonical variates is developed in terms of maximum likelihood estimators of the population. In most cases population parameters are unknown. The sample covariance matrix \mathbf{S} , is a maximum likelihood estimator of the population covariance matrix Σ . From 2.4.5, \mathbf{S} is partitioned as

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{pmatrix} \quad (2.4.8)$$

where \mathbf{S}_{11} and \mathbf{S}_{22} , are covariance matrices of \mathbf{X} and \mathbf{Y} respectively. The matrix \mathbf{S}_{12} is the covariance matrix of \mathbf{X} and \mathbf{Y} . Matrix \mathbf{S}_{21} is the transpose of \mathbf{S}_{12} .

In this regard, CCA seeks to maximise the sample correlation of two linear combinations $U = \mathbf{a}'\mathbf{X}$ and $V = \mathbf{b}'\mathbf{Y}$ by finding the first linear combination of vectors \mathbf{a} and \mathbf{b} which gives a maximum correlation. Without loss of generality, the procedure seek the maximal sample correlation

$$Corr(U, V) = \frac{\mathbf{a}'\mathbf{S}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{S}_{11}\mathbf{a}\mathbf{b}'\mathbf{S}_{22}\mathbf{b}}} \quad (2.4.9)$$

under the constraints $\mathbf{a}'\mathbf{S}_{11}\mathbf{a}=1$ and $\mathbf{b}'\mathbf{S}_{11}\mathbf{b} = 1$

This is given as

$$\max_{\mathbf{a}, \mathbf{b}} Corr(U, V) = \hat{\rho}_1^* \quad (2.4.10)$$

where $\hat{\rho}_1^*$ is the sample correlation. This is attained by

$$U_1 = \mathbf{e}'_1 \mathbf{S}_{11}^{-1/2} \mathbf{X} \quad (2.4.11)$$

$$V_1 = \mathbf{f}'_1 \mathbf{S}_{22}^{-1/2} \mathbf{Y} \quad (2.4.12)$$

and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ are eigenvalues of $\mathbf{S}_{11}^{-1/2} \mathbf{S}_{12} \mathbf{S}_{22}^{-1} \mathbf{S}_{22} \mathbf{S}_{11}^{-1/2}$ (Jeng, 2006).

Then p is assumed to be the $\min(p, q)$. In addition, the associated eigenvectors of $\mathbf{S}_{11}^{-1/2} \mathbf{S}_{12} \mathbf{S}_{22}^{-1} \mathbf{S}_{22} \mathbf{S}_{11}^{-1/2}$ are

$$\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_p \quad (2.4.13)$$

Those eigenvalues are the same for $\mathbf{S}_{22}^{-1/2} \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \mathbf{S}_{22}^{-1/2}$. However, the eigenvectors are different. The matrix $\mathbf{S}_{22}^{-1/2} \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \mathbf{S}_{22}^{-1/2}$ has eigenvectors $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_p$ with each $\mathbf{f}_i \propto \mathbf{S}_{22}^{-1} \mathbf{S}_{21} \mathbf{S}_{11}^{-1/2} \mathbf{e}_i$

The Canonical Correlation Analysis procedure repeat $\min(p, q)$ times in finding canonical correlations of linear combinations which will be orthogonal to each other. The problem of proving that vectors \mathbf{a} and \mathbf{b} maximise the sample correlation between the first pair of canonical variables U and V is done by using Jeng's(2006) proof(that is, the proof which used for the population correlation). Sample statistics are maximum

likelihood of population parameters.

Proof. Set $\mathbf{c} = \mathbf{S}_{11}^{1/2}\mathbf{a}$ and $\mathbf{d} = \mathbf{S}_{22}^{1/2}\mathbf{b}$ where the covariance matrices \mathbf{S}_{11} and \mathbf{S}_{22} are assumed to be nonsingular matrices. Then from equation (2.4.9)

$$Corr(U, V) = \frac{\mathbf{a}'\mathbf{S}_{12}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{S}_{11}\mathbf{a}\mathbf{b}'\mathbf{S}_{22}\mathbf{b}}} = \frac{\mathbf{c}'\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}\mathbf{d}}{\sqrt{\mathbf{c}'\mathbf{c}\mathbf{d}'\mathbf{d}}} \quad (2.4.14)$$

We seek to maximise $Corr(U, V) = \mathbf{a}'\mathbf{S}_{12}\mathbf{b}$. By Cauchy-Schwarz inequality the following holds

$$\mathbf{c}'\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}\mathbf{d} \leq (\mathbf{c}'\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\mathbf{c})^{1/2}(\mathbf{d}'\mathbf{d}) \quad (2.4.15)$$

$$\mathbf{c}'\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\mathbf{c} \leq \lambda_1\mathbf{c}'\mathbf{c} \quad (2.4.16)$$

where λ_1 is the greatest eigenvalue of $\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}$.

Then equality occurs for $\mathbf{c} = \mathbf{e}_1$ and

$$\mathbf{d} = \mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\mathbf{e}_1 \quad (2.4.17)$$

Now we have

$$\max_{\mathbf{a}, \mathbf{b}} Corr(U, V) = \sqrt{\lambda_1} = \hat{\rho}_1^* \quad (2.4.18)$$

with equality occurring for $\mathbf{a} = \mathbf{S}_{11}^{-1/2}\mathbf{c} = \mathbf{S}_{11}^{-1/2}\mathbf{e}_1$. Vector \mathbf{b} will be proportional to

$$\mathbf{S}_{22}^{-1/2}\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\mathbf{e}_1 = \mathbf{S}_{22}^{-1/2}\mathbf{f}_1 \quad (2.4.19)$$

$$(\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2})\mathbf{e}_1 = \lambda_1\mathbf{e}_1 \quad (2.4.20)$$

Multiply (2.4.20) both sides by $\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}$ to get

$$\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}(\mathbf{S}_{11}^{-1/2}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2})\mathbf{e}_1 = \lambda_1\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\mathbf{e}_1 \quad (2.4.21)$$

This gives

$$\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1/2}(\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\mathbf{e}_1) = \lambda_1(\mathbf{S}_{22}^{-1/2}\mathbf{S}_{21}\mathbf{S}_{11}^{-1/2}\mathbf{e}_1) \quad (2.4.22)$$

By using proportionality of \mathbf{b} , it yields

$$\mathbf{S}_{22}^{-1/2} \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \mathbf{S}_{22}^{-1/2} \mathbf{f}_1 = \lambda_1 \mathbf{f}_1 \quad (2.4.23)$$

where \mathbf{f}_1 is the normalised vector form of $\mathbf{S}_{22}^{-1/2} \mathbf{S}_{21} \mathbf{S}_{11}^{-1/2} \mathbf{e}_1$. Hence, the sign of \mathbf{f}_1 is chosen to give a positive correlation. For the first pair of linear combinations,

$$U_1 = \mathbf{e}'_1 \mathbf{S}_{11}^{-1/2} \mathbf{X} \quad (2.4.24)$$

and

$$V_1 = \mathbf{f}'_1 \mathbf{S}_{22}^{-1/2} \mathbf{Y} \quad (2.4.25)$$

Then variance of the first canonical variate of U yields the following

$$Var(U_1) = \mathbf{e}'_1 \mathbf{S}_{11}^{-1/2} \mathbf{S}_{11} \mathbf{S}_{11}^{-1/2} \mathbf{e}_1$$

$$Var(U_1) = \mathbf{e}'_1 \mathbf{e}_1$$

It simplifies to

$$Var(U_1) = 1$$

Similarly for V ,

$$Var(V_1) = \mathbf{f}'_1 \mathbf{S}_{22}^{-1/2} \mathbf{S}_{22} \mathbf{S}_{22}^{-1/2} \mathbf{f}_1$$

$$Var(V_1) = \mathbf{f}'_1 \mathbf{f}_1$$

It simplifies to

$$Var(V_1) = 1$$

where arbitrary

$$\mathbf{a}' \mathbf{X} = \mathbf{c}' \mathbf{S}_{11}^{-1/2} \mathbf{X}$$

is uncorrelated with U_1 if

$$Cov(U_1, \mathbf{c}' \mathbf{S}_{11}^{-1/2} \mathbf{X}) = \mathbf{e}'_1 \mathbf{S}_{11}^{-1/2} \mathbf{S}_{11} \mathbf{S}_{11}^{-1/2} \mathbf{c}$$

which gives a result of

$$Cov(U_1, \mathbf{c}' \mathbf{S}_{11}^{-1/2} \mathbf{X}) = 0$$

That is, $\mathbf{e}' \mathbf{c} = 0$

□

Now, it has been proven that the first pair of linear combinations, (U_1, V_1) has a maximum sample correlation of $\hat{\rho}_1^*$.

The derivation of canonical variates follows that of Geis(2012) which was used for the population covariance matrix. The sample covariance matrix was used since it is the maximum likelihood estimator of the population covariance matrix. Let (U_i, V_i) denote a pair of canonical linear combinations where $i=1,2,\dots,s \leq \min(p,q)$. It is assumed that the sample covariance matrices \mathbf{S}_{xx} and \mathbf{S}_{yy} of \mathbf{X} and \mathbf{Y} respectively, are nonsingular and invertible. The goal of CCA is to find the sample maximal correlation of

$$U_i = \mathbf{X}'\mathbf{a}_i$$

$$V_i = \mathbf{Y}'\mathbf{b}_i$$

given by

$$\hat{\rho}^* = \frac{\mathbf{a}'\mathbf{S}_{xy}\mathbf{b}}{\sqrt{\mathbf{a}'\mathbf{S}_{xx}\mathbf{a}\mathbf{b}'\mathbf{S}_{yy}\mathbf{b}}} \quad (2.4.26)$$

where \mathbf{a}_i and \mathbf{b}_i are coefficients of their respective canonical variates. Hence, the procedure seek to find vectors \mathbf{a} and \mathbf{b} that maximises

$$Corr(U, V) = \mathbf{a}'\mathbf{S}_{xy}\mathbf{b} \quad (2.4.27)$$

among all linear functions of \mathbf{X} and \mathbf{Y} . Set

$$f(\mathbf{a}, \mathbf{b}) = \mathbf{a}'\mathbf{S}_{xy}\mathbf{b} - \frac{1}{2}\sqrt{\phi}(\mathbf{a}'\mathbf{S}_{xx}\mathbf{a} - 1) - \frac{1}{2}\sqrt{\theta}(\mathbf{b}'\mathbf{S}_{yy}\mathbf{b} - 1)$$

where $\sqrt{\phi}$ and $\sqrt{\theta}$ are Lagrangian multipliers. Now taking the partial derivatives w.r.t \mathbf{a} and \mathbf{b} gives

$$\frac{\partial f(\mathbf{a}, \mathbf{b})}{\partial f(\mathbf{a})} = \mathbf{S}_{xy}\mathbf{b} - \sqrt{\phi}\mathbf{S}_{xx}\mathbf{a} \quad (2.4.28)$$

$$\frac{\partial f(\mathbf{a}, \mathbf{b})}{\partial f(\mathbf{b})} = \mathbf{S}_{yx}\mathbf{a} - \sqrt{\theta}\mathbf{S}_{yy}\mathbf{b} \quad (2.4.29)$$

Now setting these partial derivatives to zero and multiplying (2.4.28) and (2.4.29) by \mathbf{a}' and \mathbf{b}' , respectively gives the following equations

$$\mathbf{a}'\mathbf{S}_{xy}\mathbf{b} - \sqrt{\phi}\mathbf{a}'\mathbf{S}_{xx}\mathbf{a} \quad (2.4.30)$$

$$\mathbf{b}'\mathbf{S}_{yx}\mathbf{a} - \sqrt{\theta}\mathbf{b}'\mathbf{S}_{yy}\mathbf{b} \quad (2.4.31)$$

so that at maximum

$$\mathbf{a}'\mathbf{S}_{xy}\mathbf{b} = \sqrt{\phi} = \sqrt{\theta}$$

Substitute $\sqrt{\phi}$ for $\sqrt{\theta}$ and rearrange the terms. It gives the following result

$$-\sqrt{\phi}\mathbf{S}_{xx}\mathbf{a} + \mathbf{S}_{xy}\mathbf{b} = \mathbf{0} \quad (2.4.32)$$

$$\mathbf{S}_{yx}\mathbf{a} - \sqrt{\phi}\mathbf{S}_{yy}\mathbf{b} = \mathbf{0} \quad (2.4.33)$$

Multiply the former of these equations by $\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}$ and substituting it into the latter equation. This gives

$$(\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} - \sqrt{\phi}\mathbf{S}_{yy})\mathbf{b} = \mathbf{0} \quad (2.4.34)$$

or equivalently

$$(\mathbf{S}_{yy}^{-1/2}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1/2} - \sqrt{\phi}\mathbf{I})\mathbf{b} = \mathbf{0} \quad (2.4.35)$$

The polynomial of $\sqrt{\phi}$ given as

$$|\mathbf{S}_{yy}^{-1/2}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1/2} - \sqrt{\phi}\mathbf{I}| = 0$$

will give $s=\min(p,q)$ real roots. These are ordered eigenvalues $\phi_1 \geq \phi_2 \geq \dots \geq \phi_s \geq 0$ of

$$\mathbf{S}_{yy}^{-1/2}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1/2}$$

with corresponding eigenvectors $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_s$. Similarly,

$$(\mathbf{S}_{xx}^{-1/2}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1/2} - \sqrt{\phi}\mathbf{I})\mathbf{a} = \mathbf{0} \quad (2.4.36)$$

is obtained by the same procedure. It is equivalent to

$$(\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx} - \sqrt{\phi}\mathbf{I})\mathbf{a} = \mathbf{0} \quad (2.4.37)$$

Canonical variates are derived by finding vectors \mathbf{a} and \mathbf{b} which are coefficients of the following canonical variates

$$U_i = \mathbf{X}'\mathbf{a}_i$$

and

$$V_i = \mathbf{Y}'\mathbf{b}_i.$$

Multiply (2.4.37) by $\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}$ on the left to obtain

$$(\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx} - \sqrt{\phi}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx})\mathbf{a} = \mathbf{0} \quad (2.4.38)$$

Factorise $\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}$ on the right to give

$$(\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} - \sqrt{\phi}\mathbf{I})\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{a} = \mathbf{0} \quad (2.4.39)$$

which is equivalent to (0.27). Hence, \mathbf{b} is given as

$$\mathbf{b} = \mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{a} \quad (2.4.40)$$

Similarly,

$$\mathbf{a} = \mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{b}$$

is obtained by the same procedure. These vectors are coefficients of their respective canonical variables and the relation of eigenvectors shows that eigenvectors of \mathbf{a} are obtained from from eigenvectors of \mathbf{b} or vice versa. The vectors are then substituted in equation (2.4.41) and (2.4.42) in order to get the canonical variates.

Now, the i th pair of canonical variates are expressed as

$$\mathbf{a}'\mathbf{X} = (a_{i1}, a_{i2}, \dots, a_{ip}) \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \quad (2.4.41)$$

and

$$\mathbf{b}'\mathbf{Y} = (b_{i1}, b_{i2}, \dots, b_{iq}) \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{pmatrix} \quad (2.4.42)$$

where for example, \mathbf{a}' is the transpose of \mathbf{a} .

In conclusion, for large data sets the above methods are tedious and complex. This makes it difficult to do calculations by hand. To make calculations simple and easy, the analysis should be run in SPSS, SAS or in other statistical packages. However, there is a process of model building of canonical correlation analysis. CCA just like other multivariate techniques such as multiple regression analysis, it is modelled along the six stages of model-building. Aspects behind the CCA methodology is explained in the sections below.

2.4.2 Objectives of Canonical Correlation Analysis

In a research where Canonical Correlation Analysis is employed, objectives of the research should be clearly stated in line with CCA technique's objectives. The researcher should establish two theoretical datasets (the set of independent variables and the set for dependent variables). The main objectives of canonical correlation analysis are to:

- Quantify the magnitude of correlation between the two datasets;
- Deriving weights for each set of dependent and independent variables by maximising the correlation of linear combinations of each set;
- Identify the nature of relationship by measuring the contribution of each variable to the canonical functions (Hair *et al.*, 2006)

2.4.3 Designing a Canonical Correlation Analysis

Just like any other multivariate technique, CCA is affected by sample size. For reliability's sake, the requirement should be met. Small sample sizes have a tendency of not representing the correlations very well by failing to reveal meaningful relationships (Hair *et al.*, 2006). On the other hand, validity will be threatened if samples are very large by indicating statistical significance where it is not necessary. As a guideline, CCA will give meaningful results if a threshold sample sizes of at least 10 observations per variable is met (Hair *et al.*, 2006).

Due to the nature of the method of Canonical Correlation Analysis, that is, in relation to the objective of maximising correlation between the sets, the composition of each variate is of great importance(Hair *et al.*, 2006). Results are altered if any deletion or addition of a variable in either set of dependent or independent variables occurs. Hence, before applying the method of canonical correlation analysis, the researcher should have conceptually linked sets, that is, should have the correct classification of variables in both sets(Hair *et al.*, 2006). The method is also sensitive to missing data. In cases of missing data, it should be replaced by estimated values or removing cases with missing data in order to get statistically significant results(Hair *et al.*, 2006).

2.4.4 Assumptions and impact on the analysis

Canonical correlation assumes linearity among the variables just like other multivariate techniques. This will compromise results if this assumption is not met. Firstly, canonical correlation among variables is based on a linear relationship and it will not be captured by the canonical coefficient if it is non-linear(Hair *et al.*, 2006). Transformation of variables is performed if variables are non-linear. Ricketson(2008) had a different opinion on the idea of transforming variables. The argument was that even though this technique of data transformation is a suitable remedy for instances where linearity, normality and homoscedasticity are not met, the method is not universally recommended(Ricketson, 2008). This was based on the fact that the interpretation of results is based on variables in the model and transformed variables might be difficult to interpret at times(Ricketson, 2008).

The assumption of normality is not strict in the application of canonical correlation analysis since any metric variable can be accommodated in the model and large samples meet the normality assumption by the Central Limit Theorem(Thompson, 1984 and Hair *et al.*, 2006). However, normality is desirable in order to maximise correlation among the variables and required for statistical inference test for each canonical function(Hair *et al.*, 2006). All variables should be tested for normality and if this characteristic is not possessed by the variables, the transformation of variables cannot be avoided(Hair *et al.*, 2006).

Multicollinearity which is a correlation between two or more variables should be below 0.80 for CCA to reveal the relationships(Ricketson, 2008). Existence of multicollinearity among the variables will make it impossible for canonical correlation analysis to detect the impact of any variable in the canonical variate; and canonical correlation results will not be reliable(Hair *et al.*, 2006). Also, Ricketson pointed out that multicollinearity can be a problem for the correlation of $r > 0.80$ (Ricketson, 2008). Similarly, results of canonical correlation analysis are highly reliable provided that homoscedasticity is present among the variables(Hair *et al.*, 2006). However, correlation decreases if it is not present (Hair *et al.*, 2006).

In conclusion to this section, Canonical Correlation Analysis portrays all relationships provided all assumptions like normality, linearity, homoscedasticity and multicollinearity are satisfied. These assumptions should be tested when performing Canonical Correlation Analysis.

2.4.5 Deriving the canonical functions and assessing overall fit

Canonical correlation analysis entails finding canonical functions from the two sets of independent and dependent variables. Each canonical function consists of a pair of variates which separately come from the independent variables and dependent variables(Hair *et al.*, 2006). The canonical variates(functions), in the model are equivalent to the number of variables in the smallest dataset.

2.4.6 Deriving canonical variates

The procedure of deriving canonical variates is similar to that of factor analysis where the first factor to be selected is the one which accounts for maximum variance between the two sets(University of Maryland[UM], n.d). The second factor to be chosen accounts for the amount of variance which was not accounted for by the first factor and successive factors are chosen in this way(UM, n.d). Canonical variates are derived two at a time by using a method similar to factor analysis procedure. The first pair

of canonical variates to be derived is a pair which has the highest inter-correlation between the two variables from the two sets, that is, the independent variable and dependent variable sets(UM, n.d).

After the selection of the first pair, the second pair of canonical variates to be considered is the one which exhibits maximum relationship between the two sets of variables which had not been accounted for by the first pair of canonical variates(Hair *et al.*, 2006). Therefore, the process continues by considering successive pairs of canonical variates which are based on residual variance, and canonical correlation continues to decrease as each additional function is derived from the remaining variables(Hair *et al.*, 2006). In conclusion, the first pair of canonical variates displays maximum inter-relationships, the next pair exhibits second-interrelationships, and so on(Hair *et al.*, 2006).

The pairs of variates should be independent from the previous selected variates. In addition, the strength between a pair of variates, is assessed by using the magnitude of canonical correlation between them. The amount of variance in one canonical variate which has been accounted for by the other canonical variate is obtained by squaring the canonical correlation(UM, n.d). The amount of variance is also referred as the amount of shared variance between the two canonical variates(UM, n.d). Lastly, canonical roots or eigenvalues are squared canonical correlations.

2.4.7 Interpreting canonical functions

The most common method of analysing functions is to consider canonical correlation coefficients which are statistically significant above the stipulated level of significance, for example, 0.05(UM, n.d). If the functions not significant, the relationships among the variables are not interpreted(UM, n.d). Below are the three criteria which are used to decide on canonical functions to interpret(UM, n.d).

2.4.7.1 Level of significance

The most common minimum level of significance of canonical correlation is 0.05(along with the 0.01 level) and generally, this has been prompted by the availability of tables at these levels(UM, n,d). The choice of these tests is dependent on researches from different disciplines(UM, n.d). The most popular used statistic test is the F test ($p < 0.05$), which is provided by most computer packages(UM, n.d).

The evaluation of the significance of canonical roots is performed by using multivariate tests of significance such as Hotelling's trace, Pillai's trace, Wilks' lambda and Roy's gcr(greatest characteristic root criterion)(Hair *et al.*, 2006). Statistical computer packages also report(give an output) these mentioned tests for testing the general fit of the model.

2.4.7.2 Magnitude of the Canonical Relationships

The decision of which canonical functions to interpret should be based on canonical correlations(UM, n.d). There is no fixed rule regarding the sizes for canonical correlations but the decision can be based on the findings on how they contribute to the understanding of the research problem(UM, n.d).

2.4.7.3 Redundancy and Canonical Correlation Analysis

Redundancy addresses issues of strength of individual measured variables on one side of the model, that is, either that of linear combination of independent variables or dependent variables(Clark, 1975). Redundancy measures the amount of proportion of variance of the original variables of one set that can be accounted for or predicted by the variables in the other set(Clark, 1975). Redundancy index is just similar to regression R^2 and there is similarity in interpretation. The same source suggests that high redundancy may imply high ability of prediction or in other words the ability of the set of independent variables to explain variation in the other set of dependent variables(Yount, 2006; Clark, 1975).

The method of using canonical roots (squared canonical correlation) in measuring the amount of the dependent variable variance accounted for or shared with the independent variable is sometimes biased and to remedy this, the redundancy index is used instead (UM, n.d). This redundancy index is defined as the amount of variance in a canonical variate (a linear combination of dependent or independent variables) that is explained by the other canonical variate in the canonical function (UM, n.d). The following are the three steps of calculating the redundancy index (Hair *et al.*, 2006):

(a) **The amount of shared variance**

In Canonical Correlation Analysis, correlation of the dependent variate with each of the dependent variables is vital to a researcher (UM, n.d). This correlation is derived from canonical loadings which is a relationship between each input variable and its own canonical variate (UM, n.d). Each of the loadings of dependent variables is squared and thus, the amount of variation generated by each dependent variable is obtained (UM, n.d). Furthermore, the amount of shared variance is obtained by squaring the loadings and computing the average of the squared loadings.

(b) **The amount of explained variance**

The next stage is squaring of canonical correlation which is the relationship between the linear combination independent variables and linear combination of dependent variables (UM, n.d). This correlation is squared and is normally called the canonical R^2 .

(c) **The redundancy index**

The redundancy index is obtained by calculating the product of the above results, that is, $\text{Redundancy} = [\text{Mean of } (\text{loadings})^2] \times R_c^2$, where R_c is the canonical correlation (Clark, 1975). In order to get a high redundancy index, both the loadings and R_c should be high as well (Clark, 1975). Whilst redundancy indices of both the independent and dependent variates can be computed, researchers

are interested in variance extracted from the dependent set(UM, n.d). Generally, there is no fixed measure of redundancy index which can be used to interpret canonical functions, and a judgemental decision should be taken by the researcher based on the theoretical and practical significance of the research problem(UM, n.d).

2.4.8 Interpretation of the Canonical Variate

In order to interpret canonical analysis results the researcher needs to examine the canonical functions with the objective of examining the impact of each original variable in canonical correlations(UM, n.d). The ideal procedure is to examine canonical weights(standardised coefficients), canonical loadings(structure correlation) and canonical cross-loadings(Yount, 2006; UM, n.d).

2.4.8.1 Canonical Weights

To assess the importance of a variable, the sign and magnitude of canonical weights should be correctly interpreted. Variables with large canonical weights contribute more to the variates as compared to those with a small magnitude of the weight(UM, n.d). An inverse relationship is shown by the variables provided they have weights with different signs(UM, n.d). On the other hand, if the weights have the same signs, it implies that there is a direct relationship among the variables(UM, n.d). However, Yount pointed out that there are some problems in the interpretation of results, especially where weights are unstable due to multicollinearity(Yount, 2006).

According to Yount(2006), a small weight or even a negative weight might be attributed to the fact that the variance in that particular variable has already been explained by other variables. In this regard, weights failed to give reliable information about the relevance of a variable(Yount, 2006). Another problem is of instability of canonical weights from one sample to another(UM, n.d). Canonical analysis produces weights that maximise canonical correlation for any given sample of observed variables of the two sets(UM, n.d). Hence, the interpretation of results should be done with caution(UM, n.d).

2.4.8.2 Canonical loadings

Canonical loadings or canonical structure correlation are used in the assessment of contribution of each variable to its own canonical variate (Yount, 2006; UM, n.d). They measure the correlation between the variables and their respective canonical variates. Canonical loadings exhibit the variance that a particular variable shares with its own canonical variate (UM, n.d). The procedure computes correlation of a variable from one set to its variate of each independent canonical function (UM, n.d). Hence the coefficients are used to derive canonical variates where the most important variables are aligned to very large coefficients (Hair *et al.*, 2006). Just like weights, canonical loadings are affected by variability among samples and this suggests that canonical loadings can only interpret relationships in a particular sample and not to the whole population (UM, n.d). Therefore, the researcher should be cautious when using the procedure of canonical loadings so as to generalise the results (UM, n.d).

2.4.8.3 Canonical cross-loadings

From the article, canonical cross-loadings measure correlation of each variable either independent or dependent variable with the opposite canonical variate (Hair *et al.*, 2006). This procedure of cross-loadings reflects the most direct assessment of dependent and independent variable relationship (Hair *et al.*, 2006). In addition, the procedure has an advantage of expressing relationships, for example, the dependent variables with independent variates without interfering with other independent variables (Yount, 2006). Yount (2006) also pointed out that cross-loadings give a better interpretation than within-set loadings since they are less inflated.

2.4.9 Methods for interpreting the canonical variate

The main focus of this section is to compare the various methods that had been outlined above for interpreting the nature of canonical relationships. There might be a problem to some of the researchers on the choice of the best method in terms of reliability (Hair *et al.*, 2006). Even though canonical weights perform very well in a multivariate setting, they can produce unreliable results since they are sensitive to multicollinearity (Hair *et al.*, 2006). In comparison, the canonical-loadings procedure

is somewhat more appropriate than canonical weights(Hair *et al.*, 2006). The remaining procedure, that is, canonical cross-loadings is based on the principle of linearity and is the most preferred method of them all because the procedure can transform a canonical model to a single latent construct(Hair *et al.*, 2006). Furthermore, the researcher should compute canonical cross-loadings manually since most computer programs do not provide cross-loadings or to rely on canonical loadings(Hair *et al.*, 2006).

2.4.10 Limitations of the canonical correlation procedure

The following limitations were noted when performing canonical correlation analysis procedure(Hair *et al.*, 2006):

- (i) Canonical correlation does not measure variance of individual variables but it only measures shared variance of a set of variates.
- (ii) The researcher should be aware of canonical weights since they are unstable.
- (iii) Canonical weights are derived to maximise the correlation between linear composites, not the variance extracted.
- (iv) Sometimes canonical variates can be difficult to interpret.
- (v) Identification of relationships between the subsets of independent and dependent variables can be difficult because of unavailability of precise statistics to interpret canonical correlation.

The reliability of results depends on the appropriateness of the research methodology and making use of the assumptions and correcting the applicability of the canonical correlation procedure.

2.5 Concluding Remarks

The literature has reviewed the underlying theories about crime where aggravated robbery is one type of crime. Despite the differences in the causation of crime, some theorists agreed at some point. Prevention and intervention of aggravated robbery was suggested to be mainly embedded into these theories like the General Strain Theorem. Canonical correlation analysis methodology which is used to reveal interrelationships between two datasets has been discussed at length above. The next section describes the methodology which was used in the research.

Chapter 3

Research Methodology

3.1 Chapter outline

The previous sections have focused on the philosophical aspects of poverty and aggravated robbery crime, and the theoretical aspects of Canonical Correlation Analysis. This section reviews the research methods and design which were used in the research, geographical setting of the research, a discussion of the population and sample, instrumentation, data collection methods and data analysis or analysis plan. In addition, this section outlines how quantitative research methodology was employed and an application of a multivariate technique in analysing the correlation between poverty and aggravated robbery crime. This technique is known as Canonical Correlation Analysis and it was used to analyse the relationship between poverty and aggravated robbery crime in Limpopo Province for a period of two and half years.

In other words, this chapter outlines how the research problem was solved through answering research questions. As stated in Section 1, the purpose of the study was to find out the relationship between poverty and aggravated robbery in Limpopo Province. In conclusion to this introductory section, the above outlined aspects are discussed in the sub-headings below.

3.2 Research Methodology and Design

The main objective of the research design is to maximise reliability by reducing bias as much as possible. Research design is a conceptual structure which comprise collection, measurement and analysis of data(Kothari, 2004). Furthermore, it should have the following features such as, (i)clear statement of the research problem; (ii)procedures and techniques to be used; (iii)the population to be studied; and (iv) approach for data collection and data analysis(Kothari, 2004). There are several types of research designs. According to Creswell(2003), there are three types of framework design, namely, quantitative, qualitative and mixed methods approaches. In this research, quantitative research has been adopted according to the nature of the data.

3.2.1 Quantitative research

There are different approaches to the definition of quantitative research. Nykiel(2007) defined it as a way of quantifying relationships between variables with an objective of establishing facts, making predictions and testing of hypotheses by collecting measurable information that can be tracked over time. This procedure will enable the researcher to make inferences about causality between variables or datasets. Based on this approach, all aspects such as testing of hypotheses and determination of relationships between variables will be described in the sections below.

In addition, Creswell defined quantitative research as a research that employs strategies of inquiry such as surveys, collects data on predetermined instruments that yield statistical data in order to explain phenomena(Creswell, 2003; Sukamolson, n.d). With this in mind, quantitative research can be categorised into different types such as, descriptive research, experimental research, quasi-experiment research, correlational research, survey research, casual-comparative research and inferential research(it explains a phenomena, for example, relationship between variables and it uses inferential statistics such as multiple regression and canonical correlation analysis)(Nykiel, 2007; Sukamolson, n.d). In this research, correlational research design was used with the intention of determining the extent to which poverty and aggra-

vated robbery are related in Limpopo Province.

3.2.1.1 Rationale of the research design

According to the nature of the research(it deals with secondary data which is in the form of numbers), that is, the nature of hypotheses and objective of the research, quantitative approach was chosen as an optimum choice mainly for the following reasons:

- Numbers enable greater precision in measurement. There is a well-developed theory of reliability and validity to assess measurement errors and this will enable researchers to know how much confidence to place in their measures. Trends can be easily analysed, for example, the trend of aggravated robbery in Limpopo Province.
- Quantitative measurements facilitate comparison. They allow researchers to get the reactions of many people to specific stimuli and to compare responses across individuals. For example, in canonical correlation analysis this comparison is employed by assessing the relationships of canonical covariates of the two sets of variables. In this research quantitative measurements was of great benefit since it enabled the comparison between types of aggravated robbery and also among variables of poverty. Hence, the hypothesis(see Section 1.7) was tested adequately by using the above-mentioned approach.
- There are well-established statistical methods for analysing data such as canonical correlation analysis, multiple regression analysis, and so on. Data can easily be summarised. Reliable results pertaining to the population will be obtained provided data was collected by using reliable methods.
- Quantitative measurements fit well with hypothetic-deductive approaches. Hypothesised relationship between variables can be specified using a mathematical model, and the methods of statistical inference can be used to see how well the data fit the predictions. For example, hypothesis on Section 1.7 was tested by using inferential statistics such as canonical correlation analysis.
- Sampling theory can be used to estimate how well the findings generalise beyond the sample in the study to the wider population from which the sample was drawn(Barker *et al.*, 2002).

3.3 Variables used in the research

Sekaran and Bougie(2010) defined a variable as anything that can assume differing or varying values. These values can differ as time passes for the same object or person(Sekaran and Bougie, 2010). Variables are classified as independent variable(predictor variable) and dependent variable(the response variable). The independent variable come first to the dependent variable.

3.3.1 Independent variables

In addition, Wimmer and Dominick(2011) outlined that variables are classified in terms of their relationship with one another. According to Wimmer and Dominick(2011), the distinction between independent and dependent variables lies mainly on the purpose of the research. In this research, the independent variables, that is, levels of poverty are categorised into gender(GENDER), marital status(MSTAT), educational attainment(EDUA), employment status(EMPS), age(AGE) and race(RACE). These variables are justified by using literature on determinants of poverty. The correlation of these variables had been investigated in other previous research.

Anyanwu(2012) argued that poverty increases as one grows older. Hence from the age perspective, productivity decreases as age increases and an individual will need more money to compensate for lost productivity and income. This justifies age as a variable of poverty in this research. Also, educational attainment correlates with poverty. Literature revealed that labour is an asset of the poor since education increases labour productivity(Anyanwu, 2012). This implies that there will be a significant decrease of poverty if poor people are educated since the economy will benefit in their labour(Anyanwu, 2012).

Furthermore, gender and race are also variables of poverty(Jordan, 2004). Race is regarded as factor of poverty since economic disparities in the United States of America in 1947 were due to race(Jordan, 2004). Even in post independent South Africa, poverty is still based on racial lines. Poverty is more pronounced among Blacks than

among other races due to discrimination during the apartheid era. Anyanwu(2012) also revealed convincing evidence about gender as a poverty variable.

Jordan(2004) pointed out unemployment as a structural economic factor of poverty. Generally, work is regarded as a mechanism which uplifts people from poverty. In this research, employment is a dichotomous variable(employed and not employed). An employed individual will be able to get his or her basic needs(the individual will be operating above the poverty line or threshold). Marital status has been classified as married or not married(single). Anyanwu(2012) argued that married couples are relieved from hardships since they source assistance from relatives. Also, a marital relationship may mean a relief to poverty since this will limit economic hardships(Anyanwu, 2012). For example, male-headed households have a higher income base since there will be more earning members(Anyanwu, 2012)

3.3.2 Dependent variables

The dependent variable is the number of aggravated robbery crimes which is an observed value, and its value is presumed to be dependent on the influence of independent variables(Wimmer and Dominick, 2011). Crime levels are categorised into truck hijacking(TRUH), house robbery(HROB), business robbery(BROB), cash-in-transit robbery(CIT), carjacking(CARJ), bank robbery(BAROB) and street/common robbery(CROB). Then the two groups of variables, poverty and aggravated robbery crime are expressed as linear combinations of their respective levels(Ho, 2010). The linear combinations for the i th pair of canonical variables are calculated as(Manly, 1986)

$$Poverty_i = \left(a_{i1}, a_{i2}, \dots, a_{i6} \right) \begin{pmatrix} GENDER \\ EMPS \\ MSTAT \\ RACE \\ AGE \\ EDUCA \end{pmatrix} \quad (3.3.1)$$

and

$$Crime_i = \begin{pmatrix} HROB \\ BAROB \\ CROB \\ CARJ \\ TRUH \\ CIT \\ BROB \end{pmatrix} \quad (3.3.2)$$

where from (2.4.37) and (2.4.38) $\mathbf{a}'_i = (a_{i1}, a_{i2}, \dots, a_{i6})$ which is a vector of corresponding eigenvectors to the latent roots of $|\mathbf{S}_{xx}^{-1} \mathbf{S}_{xy} \mathbf{S}_{yy}^{-1} \mathbf{S}_{yx} - \lambda \mathbf{I}| = 0$. These eigenvectors are coefficients of poverty variables. Similarly, $\mathbf{b}'_i = (b_{i1}, b_{i2}, \dots, b_{i7})$ are eigenvectors which are also coefficients of the crime variables (aggravated robbery crime variables).

Finally, the focus was on the comparison of levels of crime (dependent variables) compared with levels of poverty. Each of the levels was compared against one another.

3.4 Research Questions and Hypothesis

3.4.1 Research questions

In order to develop an in-depth analysis of the relationship of poverty and aggravated robbery crime in Limpopo Province, statistical inferences were used to research questions below. Research questions were formulated. Hence the research questions are stated below as:

- Is there any meaningful relationship between poverty and aggravated robbery crime in Limpopo Province?

The sub questions used to answer the main research question above are:

- Is there any correlation between poverty criteria and aggravated robbery crime criteria?
- In a set of variables of aggravated robbery, which variable has the most and which one has the least impact on creating a meaningful relationship between poverty and

aggravated robbery crime?

- In a set of variables of poverty, which variable has the most and which one has the least impact on creating a meaningful relationship between and poverty and aggravated robbery crime?

3.4.2 Research hypothesis

The research was more inclined to hypotheses-testing research with the hypothesis stated below as

$$H_0: \rho_1 = \rho_2 = \dots = \rho_s = 0$$

$$H_1: \rho_i \neq 0 \text{ at least one } i=1; 2; \dots; s$$

and $s = \min(p, q)$ where p and q are variables of the two datasets.

This hypothesis was tested by using canonical correlation analysis. A software package known as SPSS was used to run the analysis. In addition, the research design tested the strength of relationships as well as direction (commenting on positivity or negativity) of relationships.

3.5 Area of the Study

The research was conducted in Limpopo Province which shares borders with Botswana, Zimbabwe and Mozambique. Limpopo province is one of the nine provinces of South Africa. It has five districts, namely, Vhembe, Mopani, Capricorn, Waterberg and Greater Sekhukhune. The map of Limpopo province is shown on Figure 3.5.1 below. This map shows the area of study as well as sampled districts, that is, Capricorn, Vhembe and Mopani.

Limpopo Province covers a size of $123\,910\text{km}^2$ which is 10.2% of South Africa's land area (Kyei and Gyekye, 2011). According to StatsSA (2012), the National Census of 2011 showed that Limpopo Province had different population composition where the greatest percentage was found among Blacks. In other words, it is a multicultural society with four subgroups, namely, African, Coloured, Asian/Indian and White (Punt *et al.*, 2005). From the report on the findings of the National Census of 2011, Blacks

constitute 96.8%, Coloureds 0.3%, Indian/Asian 0.3% and Whites 2.6% (StatsSA, 2012). Punt *et al.*,(2005) outlined that the province is marred by high poverty rates and inequalities in the distribution of income between various subgroups. Majority of the population lives in rural areas.

The unemployment rate in 2011 was pegged at 39.5% and femaleheaded families were at 50.4% (StatsSA, 2012). The report also gave marital status statistics for 2011. Statistics for married/living together, never married, widowed/widower and divorced/separated were 25.4%, 69.0%, 4.3% and 1.3% respectively(StatsSA, 2012).

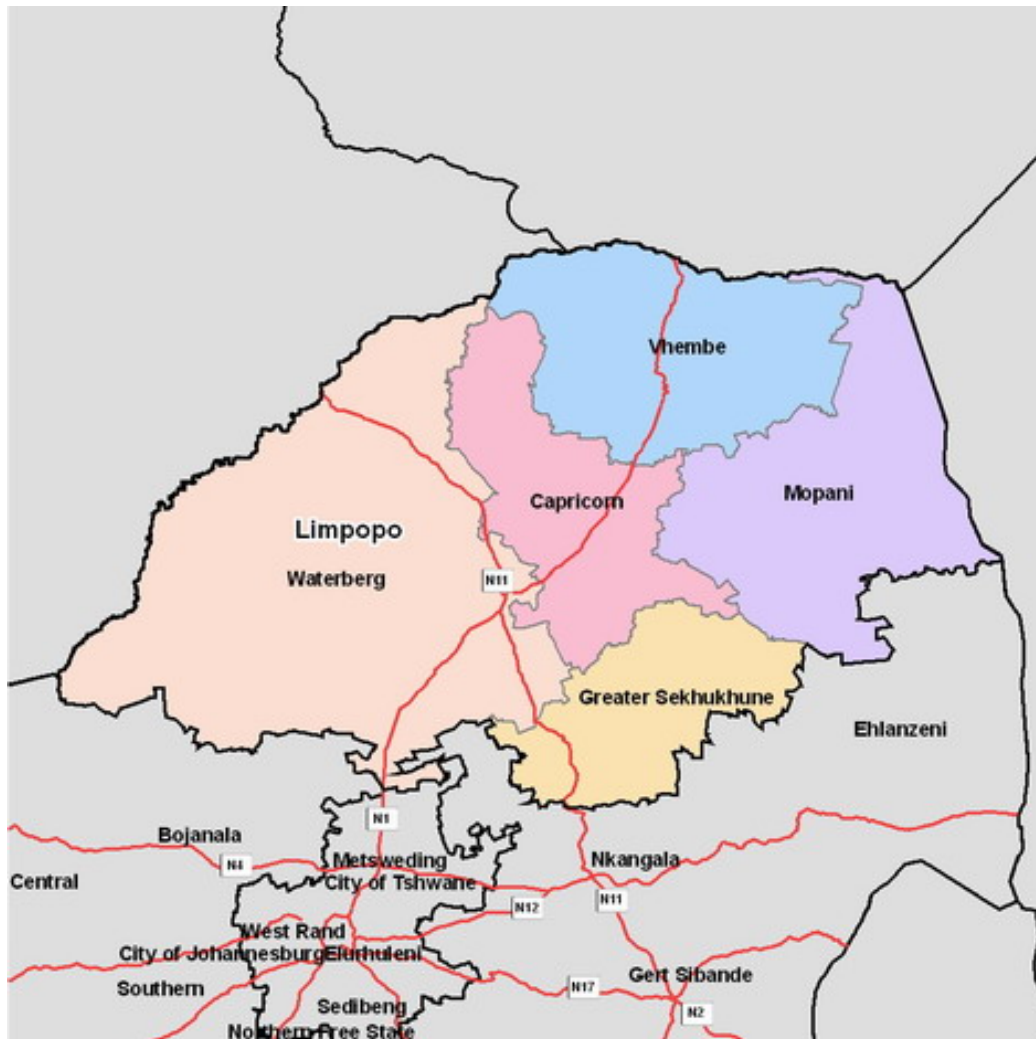


Figure 3.5.1: Limpopo Province Districts Map

Adapted from PROVIDE Project Background Paper 2009: 1(9)

3.6 Pilot Study

The first phase of the research was constituted by a pilot study which made use of police stations in Thohoyandou Cluster. The pilot study was based on Thabane *et al.*,(2010)'s argument about the main purpose of a pilot study. Thabane *et al.*, (2010) defined a pilot study as a test of the feasibility of the proposed research methods and procedures for a later use or to be used on a large scale. The main specific objectives of the study were to assess how data of convicted offenders of aggravated robbery crime were recorded(process of analysing these records is known as docket analysis); to assess time and budget for the final study(Thabane *et al.*, 2010); to determine sampling methods to be employed during data collection; and to determine the feasibility of the research and procedures to be used on a large scale. The study findings showed the feasibility of the study with variables similar to those that appear in dockets (criminal records). A recording sheet of data was designed which conforms to SPSS data template or spreadsheet. Due to the sensitivity of the criminal data, SAPS officials' advice was to use a different coding system of the accused offenders or criminals rather than using CAS(Crime Administration Systems) coding. A program known as Business Intelligence System(BIS) was used to access data of aggravated robbery crime from SAPS data archives.

3.7 Target Population

The credibility of the study depended on the proper selection of the target population, choice of sampling design and execution of the design(Henry, 1990). According to Henry(1990), target population is defined as a group about which the research seeks to make some generalisations. The selection of target population is mainly based on objectives or purpose of the study(Henry, 1990). From this point of view, the research proposed a canonical correlation analysis procedure in assessing the relationship of poverty and aggravated robbery crime in Limpopo Province. The target population comprised of data of aggravated robbery offenders from 1 January 2011 up to 30 June 2013 in Limpopo Province. These offenders should have been jailed for the offence and those who were acquitted were not considered.

3.8 Sampling Design

Sampling design involves the selection method employed to choose the sampling units and determining the sample size(Henry, 1990). Sampling methods were employed in selecting a sample instead of the entire population of convicted aggravated robbery crime offenders in Limpopo SAPS records. This was due to financial constraints in data collection, time framework of the research, costs of analysis of large data, follow-up evaluations during the collection of data and the administration of personnel during data collection(Henry, 1990). Sampling guarantee quality of results. The following sections are devoted to explaining the sampling technique and the size of the sample which was used in this research project.

3.8.1 Sampling methodology

Due to the nature of the study, multistage sampling was used. It is one type of probability sampling. This method enhances the credibility of the study. Probability sampling encompasses a known non-zero probability for each sampling unit of being included in the sample(Henry, 1990). This removed bias in selection of convicted aggravated robbery crime offenders from the sampling frame which was a list of aggravated robbery criminals from 1 January 2011 to 30 June 2013.

A multistage sampling was used in this research due to financial constraints and other factors such as administration of collection of data. It refers to sampling in the stages where in this research a two-stage sampling was employed(Henry, 1990). Limpopo Province has five districts as stated earlier in some sections above. These districts were viewed as primary sampling units(PSUs) which were sampled first by using Simple Random Sampling(Henry, 1990). Henry(1990) stated the probability of the selection of each sampling unit as:

$$p = \frac{n}{N} \tag{3.8.1}$$

where p is the probability of selection,
 n is sample size, and

N is number of units in the population.

Hence, the sample of districts consists of three districts and the probability of selecting a district was 0.6. These districts were coded 1-5: Waterberg, Capricorn, Sekhukhune, Vhembe and Mopani. Simple random sampling was used to select three districts. A calculator was used to generate random numbers. Mopani, Capricorn and Vhembe districts were chosen. Each district has three clusters of police stations and data was collected from these clusters.

After the selection of districts, data was collected from the nine clusters of the three districts. Police stations of each district were grouped into three clusters. Members or units of the sample were then selected from each police station in each cluster using simple random sampling so as to increase precision. Simple random sampling was used in order to get rid of the subjectivity judgement during the selection of aggravated robbery crime offenders from the population of offenders of this type of crime(Henry, 1990). Roughly, it was expected to sample ten to fifteen offenders from each police station based on the sample size(It is explained in the section below).

The officers in the Criminal Investigation Department(CID) used the BIS program to access details of offenders using their CAS numbers. The system(BIS program) has a numerical list of offenders per type of aggravated robbery crimes. Hence, it was easier to use random numbers which were generated by a calculator in the selection of members of the sample.

In summary, the process of sampling was two-stage in the sense that districts were selected first and then offenders were chosen from police stations of the selected districts. In this design, threats to validity and reliability were eliminated(Henry, 1990). Another aspect which was considered regarding the validity of the study was sample size. If this aspect is overlooked it contributes to Type II error which occurs when a null hypothesis is not rejected when it is false(Henry, 1990).

3.8.2 Sample size

In any research design, sample size plays a crucial role. It enables one to make inferences about the population under study by using sample statistics where population parameters cannot be analysed. Henry(1990) used a diagrammatic representation to show the relationship of sample size and sampling variability which has an inverse relationship. An increase of sample size implies that there will be a decrease in sampling error which translates to a small difference of sample results to those of the population(Henry, 1990). Sometimes it is not necessary to have large samples since this will be a waste of time and even resources which should have been put to other uses. As the sample size increases, it reaches a maximum point where results will never improve except for a census(Krejcie and Morgan, 1970). This argument is supported by Figure 3.8.1 below.

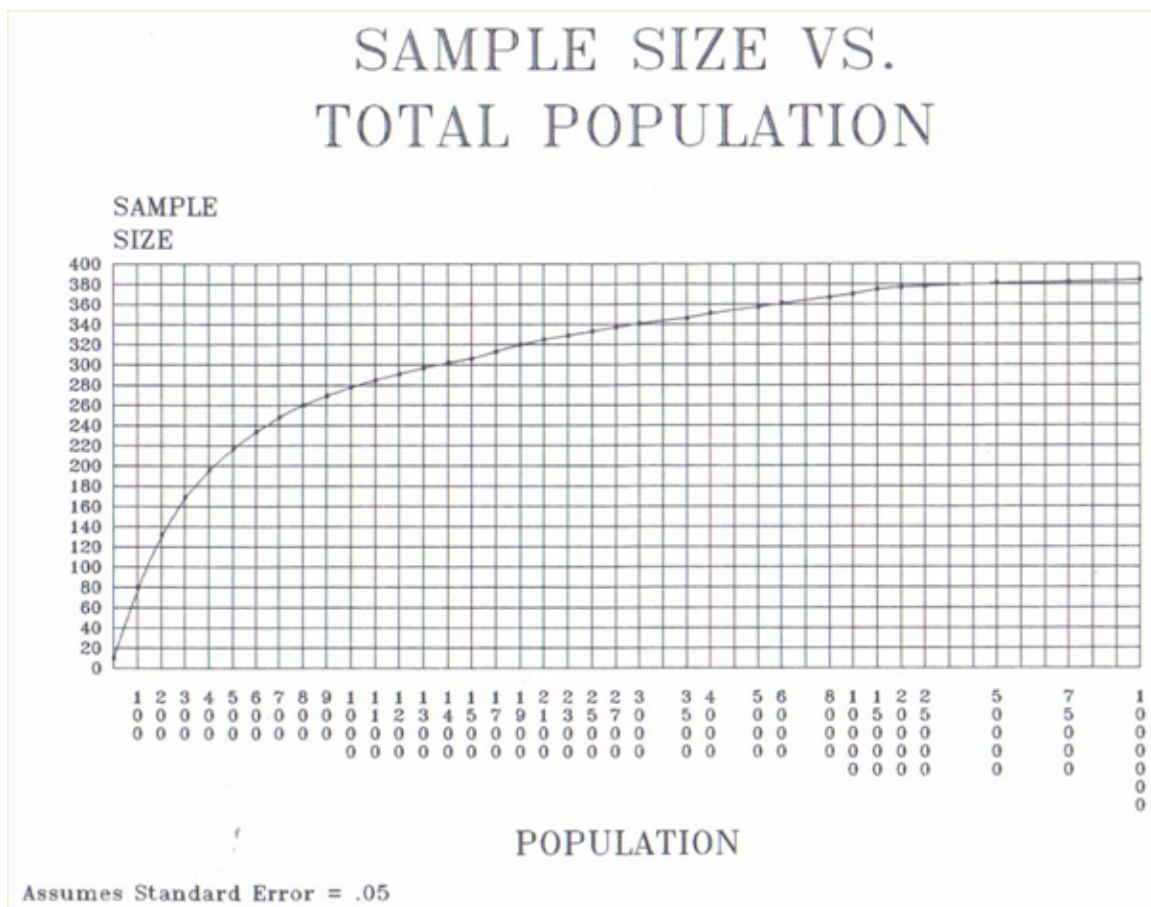


Figure 3.8.1: Comparison of Sample size and Population

Adapted from Krejcie and Morgan(1970)

Having outlined the rationale of a sample size, the sample size was determined by using the following principles. Determination of the sample size was based on population size, sampling error to be tolerated, study objectives and the risk of selecting an inappropriate sample (Israel, 1992; Chow, Shao and Wang, 2003). The sample was drawn from a list of aggravated robbery criminals in Limpopo Province as from 1 January 2011 up to 30 June 2013. The overall statistics for the whole province is shown in Table 3.8.1 with particular reference to types of aggravated robbery crime for that given time frame.

Table 3.8.1: Limpopo aggravated robbery crime statistics: 01/01/2011 – 30/06/2013

Type of crime	Number of crimes recorded
House robbery	1 812
Bank robbery	0
Street robbery/Common robbery	8 426
Carjacking	458
Truckjacking	54
Cash-in transit	16
Business robbery	2 799
TOTAL	13 565

3.8.3 Sample size calculation

There are several approaches for determining sample size such as tables and formulas. Calculations of the sample size was based on precision analysis. Precision analysis refers to calculation of the sample size by controlling *type I* error with probability α (Chow, Shao and Wang, 2003). However, the probabilities of *type I* error and *type II* error are given as

$$\begin{aligned}\alpha &= P(\text{type I error}) \\ &= P(\text{reject } H_0 \mid H_0 \text{ is true})\end{aligned}$$

$$\begin{aligned}\beta &= P(\text{type II error}) \\ &= P(\text{fail to reject } H_0 | H_0 \text{ is false})\end{aligned}$$

respectively. In precision analysis, calculation of the sample size is based on the upper half of the $(1 - \alpha)100\%$ confidence interval (Chow, Shao and Wang, 2003). A narrow confidence interval increases the precision in estimating an unknown parameter (Chow, Shao and Wang, 2003). Then the maximum half width is called the *maximum error*. Now, Krejcie and Morgan's formula was based on simple random sampling without replacement of sampling units from a finite population. The maximum half width of $(1 - \alpha)100\%$ confidence interval of estimating the population proportion p parameter with d as the *maximum error* or margin of error which one is willing to accept is given by (Chow, Shao and Wang, 2003)

$$d = Z_{\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n} \left(1 - \frac{n-1}{N-1}\right)} \quad (3.8.2)$$

From (3.8.2), $Z_{\frac{\alpha}{2}}$ is replaced by $\sqrt{\chi_{\alpha}^2(df)}$ (Chow, Shao and Wang, 2003). The derivation of the formula is as follows

$$\begin{aligned}d &= \sqrt{\chi_{\alpha}^2} \sqrt{\frac{P(1-P)}{n} \left(1 - \frac{n-1}{N-1}\right)} \\ d^2 &= \chi_{\alpha}^2 \left[\frac{P(1-P)}{n} \left(\frac{N-n}{N-1}\right) \right] \\ d^2(nN - n) &= \chi_{\alpha}^2 P(1-P)N - \chi_{\alpha}^2 P(1-P)n\end{aligned}$$

Grouping terms with n and applying factorisation will yield the following result

$$\begin{aligned}nNd^2 - d^2n + \chi_{\alpha}^2 P(1-P)n &= \chi_{\alpha}^2 NP(1-P) \\ n(d^2(N-1) + \chi_{\alpha}^2 P(1-P)) &= \chi_{\alpha}^2 NP(1-P)\end{aligned}$$

Thus

$$n = \frac{\chi_{\alpha}^2 NP(1-P)}{d^2(N-1) + \chi_{\alpha}^2 P(1-P)}$$

is used to compute the minimal sample size.

Hence, the study used the above formula mainly due to easy flexibility in terms of population. Tables have fixed population sizes (N). The above formula popularly known

as the Krejcie and Morgan's(1970) formula, was used to estimate an appropriate minimal sample of aggravated robbery criminals. Krejcie and Morgan's formula(1970) is as follows:

$$n' = \frac{\chi^2 NP(1 - P)}{d^2(N - 1) + \chi^2 P(1 - P)} \quad (3.8.3)$$

Where

n' = sample size

χ^2 = table value of chi-square for 1 degree of freedom at the desired confidence level (3.841)

N = The population size

P = The population proportion (assumed to be 0.50 since this would provide the maximum sample size).

d = The degree of accuracy expressed as a proportion (0.05).

Hence the minimum sample was calculated as:

$$n' = \frac{3.841 \times 13\,565 \times 0.50 \times 0.50}{0.05^2 \times 13\,564 + 3.841 \times 0.50 \times 0.50} = 374 \quad (3.8.4)$$

to the nearest unit.

Cochran's(1977) formula is an alternative method of calculating the sample size. The formula used for categorical data is stated below. Just like the Krejcie and Morgan's formula(1970), determination of the sample size is also based on the desired limits of error. Basically this is the amount of error required to tolerate when estimating sample estimates(Cochran, 1977). Application of Cochran's(1977) formula assumed simple random sampling of selection of the sampling units and p the proportion parameter to be normally distributed. The formula that connect the margin of error, that is, desired degree of precision in estimating sample estimates is

$$d = t \sqrt{\frac{N - n}{N - 1} \left(\frac{PQ}{n} \right)} \quad (3.8.5)$$

Where t =the abscissa of the normal curve that cuts an error of α of the tails,

d = acceptable margin of error for proportion that can be incurred at probability α ,

N = the population size,

P = the population with a specific characteristic,

Q = the true proportion without a specific characteristic.

After solving for n , Cochran's(1977) formula is given by

$$n = \frac{\frac{t^2 PQ}{d^2}}{1 + \frac{1}{N} \left(\frac{t^2 PQ}{d^2} - 1 \right)} \quad (3.8.6)$$

Usually the t -value for alpha level of .05 for sample size of more than 120 is 1.96 and for smaller samples of 60, it will be 2.00(Cochran, 1977). From Cochran(1977), for large N , the approximation is given as

$$n_0 = \frac{t^2 pq}{d^2} \quad (3.8.7)$$

where p is an estimate of a population proportion parameter P .

Then the sample size, n is given by

$$n = \frac{n_0}{1 + \left(\frac{n_0 - 1}{N} \right)} \quad (3.8.8)$$

Application of Cochran's formula to determine the sample size

$d = 0.05$, $p = 0.5$, $\alpha = 0.05$, $t = 1.96$ From (3.8.7)

$$\begin{aligned} n_0 &= \frac{(1.96^2)(0.5)(0.5)}{0.05^2} \\ &= 384 \end{aligned}$$

During the period of the study, 13 565 offenders were convicted of aggravated robbery crime. Thus the minimal sample size according to Cochran(1977) was computed and the value of n was

$$\begin{aligned} n &= \frac{384}{1 + \left(\frac{383}{13565} \right)} \\ &= 373 \end{aligned}$$

The computed sample size from Cochran's formula was almost equal to the one obtained from Krejcie and Morgan's(1970) formula which was 374. This proved reliability of the minimal sample of the research design. However, an analysis which involved multiple regressions, analysis of covariance, log linear analysis as well as canonical correlation analysis needs large samples such as 200-500(Israel, 1992). Hence it was increased to 559.

3.9 Data Collection

Research data which was collected and analysed during the research was secondary data. Collins(2010) defined secondary data as data which is collected by someone specifically not for the user(second researcher) and under the conditions not known to the user. By using this principle, the data for aggravated robbery crime was collected from three districts in the province. Authority was sought before the commencement of data collection. The research used data from organisational records at the CID Department(crime detectives). Hence correspondences to the Provincial Police Commissioner were done in time. For the analysis of poverty levels in Limpopo Province, variables such as gender, age, employment status, marital status, race and level of education were used.

Secondary research has some advantages. One of the advantages is that it is cheaper and more quickly available than primary data. The time frame for data collection is short and it provides easy access to knowledge(Collins, 2010). From a financial point of view, costs for acquiring are very low. Also secondary data can be used for analysis of a research on a large scale(Collins, 2010).

Disadvantages adapted from Collins(2010) are as follows:

- ❖ Secondary data is not presented in a form that is required by the researcher.
- ❖ It can also fail to provide a full version of data.
- ❖ Time can also be a factor; data collected 5 years ago is different from that which is collected now.
- ❖ It is largely selfgoverned and must therefore be scrutinised.
- ❖ Data may be difficult to access or expensive to obtain.

Having said much on advantages and disadvantages of secondary data, this type of data can be categorised into documentary, survey and multiple-source data(data set which has been combined before another analysis)(Collins, 2010). Government can

generate secondary data such as crime data and census data. When dealing with secondary data, validity and reliability of the data should be verified.

3.10 Data Coding

From the nature of the data, the format didn't conform to the data format of CCA since it is designed for continuous data. Without data coding it was going to be impossible to capture the correlations between poverty and aggravated robbery crime. Even computer programs such as SPSS, cannot read unconverted data. Canonical Correlation Analysis methodology is employed to variables which are in interval/ratio scale. Poverty variables were dichotomous variables. For example, gender was coded as "0" for female and "1" for male. Due to space constraint, other data codings are presented in Appendix C. A justification can be given to why age was taken as a categorical variable. This was based on Schmallegger's(2009) and other researchers' findings that hormonal decrease of testosterone which is linked to violence and aggression, start to decline at 25 years. Also, the objective was on the relationship of age to other variables and to the two datasets. Therefore, age variable was taken as a zero-one scale("0" for 24 years and below and "1" for above 24 years). Independent variables were now numeric and this forced CCA to reveal interrelationships between the two datasets. By so doing, the objective was achieved.

However, hypothesis testing in the context of General Linear Multivariate Model(GLMM) where CCA, Principal Component Analysis, Multivariate analysis of variance(MANOVA), and so on, requires fulfilment of multivariate assumption(Generalized Linear Models[GLM], 2002). Hence, the absence of multivariate normality of the response variable which belongs to count data, for example, will lead into Type II error. Count data take only non-negative integer values{ 0, 1, 2, 3,...} and they arise from counting(GLM, 2002). Count data are often Poisson distributed. Since crime data is an example of count data, it implies that it follows a Poisson distribution and do not fulfill the normality multivariate assumption. This necessitate the transformation of the response variable(GLM, 2002). Hence, response data is normalised since this

removes heterogeneity of variance. GLM(2002) suggested transformation of data by using logarithmic or square root transformation as a means of stabilising the sample variance of the response variable. The coding was “0” for not committing aggravated robbery crime and “1” for committing aggravated robbery crime(Appendix C).

As a requirement of SPSS program, districts were coded as “1” for Vhembe district, “2” for Mopani district and “3” for Capricorn district. All the nine clusters in these districts were coded. For example, “1” for Musina cluster and “9” for Lebowakgomo cluster. A full list list of these codings are in Appendix C.

3.11 Reliability and Validity

The research focused on data from official crime statistics(quantitative data collected by government) which were obtained from the South African Police Service(SAPS). As a rule of thumb, reliability and validity of official data should be examined. It is discussed under the following subheadings.

3.11.1 Reliability

Golafshani(2003) defined reliability as:

“ The extent to which results are consistent over time and an accurate representation of the population under study is referred to as reliability. If the results of a study can be reproduced under a similar methodology then the research instrument is reliable”. The key elements of reliability raised are consistency and accuracy of the measurement instrument. The same conditions subjected to all subjects under study, should be the same so as to minimise reliability errors. Such errors were minimised by assisting in the transcription of data by the researcher. This was done to prevent data manipulation since crime statistics is sensitive(it is released only to authorised people) and data can be altered for reasons best known to the organisation officials. In addition to these precautionary measures, official crime statistics data is regarded as a reliable indicator(Burger *et al.*, 2009). Since there is no tangible evidence that crime data have been manipulated, this made the statistics the only official indica-

tors which showed trends in crime in South Africa(Burger *et al.*, 2009). Hence, official crime data was regarded as reliable.

3.11.2 Validity

Besides the threat of reliability to the research study, validity is another aspect which any researcher considers when carrying out research. According to Golafshani(2003), validity in quantitative research should determine the ability of the research to truly measure what it is intended to measure or how truthful the research results are. Validity has two essential components, that is, internal and external validity(Handley, n.d). Internal validity encompasses the legitimacy of the results of the study based on the selection procedure of subjects in the sample. It also justify the methodology of data collection and analysis of results. In this study, multistage sampling was used with sample units having the same probability of being selected. The two key concepts of validity and reliability were kept on check so as to achieve the objective of the research by answering the research questions. External validity which entails the generalisability of the research results, was of paramount importance during the study. The objective was to generalise the research results of aggravated robbery crime to the population of aggravated robbery criminals in Limpopo Province.

3.12 Data Analysis and Interpretation

Inferential data analysis of canonical correlation determined the complex interrelationships between poverty and aggravated robbery crime data.The Statistical Package for Social Sciences(SPSS) was used for data analysis of this research. Inferential statistics such as canonical correlation analysis was very useful in investigating the relationship between aggravated robbery crime and poverty in Limpopo Province. The statistical tests on hypotheses about the combination(s) of the criterion(aggravated robbery) variables are related to combination(s) of predictors(poverty) made use of canonical weights(standardised coefficients), canonical loadings, canonical cross-loadings, redundancy coefficient and canonical correlation. These were employed to interpret the canonical analysis results and to answer the research question: “Is

there any meaningful relationship between aggravated robbery crime and poverty in Limpopo Province?”. In doing so, this answered the hypothesis in Section 1.7.

Hence the magnitude of the canonical correlation was determined whether there was a relationship between poverty and crime in Limpopo or not. In the analysis, canonical cross-loadings were used. For example, they were used to determine the main poverty variables (levels) that contributed to aggravated crime or vice versa. From Section 2.4.8.2, canonical cross-loadings measure the correlation of a variable in either independent or dependent variable with the opposite canonical variate. Hence, the comprehensive analysis and interpretation was based on computer outputs.

Finally, it was planned to disseminate the research findings to the academic community through an accredited journal in statistics. Government departments and all stakeholders would have access to the journal article.

3.13 Research Ethics

Penslar(1995) defined research ethics as the philosophical study of normative behaviour, that is, the expectations of morals by the researcher. In other words, research ethics entails providing solutions to specific problems that could arise in the process of conducting the research(Penslar, 1995). Just like any other research, research ethics should be considered. Confidentiality is of paramount importance. In this research, offenders were coded instead of using their real names.

In Section 3.9 it had been outlined that some of the researches incorporates secondary data where the researcher used data which had been collected by another researcher. For a study which employs secondary data, a full ethics application is not required, that is, ethics certificates and the involvement of research committee(University of Pretoria, 2003). According to the document of the University of Pretoria, a letter should be written to the owners of the data in order to seek permission to access

data(University of Pretoria, 2003). Hence, correspondence letters with SAPS Provincial Commissioner Headquarters of Limpopo Province were written(see Appendix B).

Supporting letters were also written by the supervisor. These letters sought permission to access data on aggravated robbery crime from police stations in Limpopo Province. In order to convince those who granted authorisation to access of data, the research topic and main objective(s) of the research were included in the application letters. A draft proposal was attached as a supporting document to the application of data access. The draft explained the validity, reliability or trustworthiness, research questions, research objectives and the research's contribution to all stakeholders(University of Pretoria, 2013). However, application for authorisation was done before the research topic was in its final version. By then, the topic was as follows: "A quantitative research of aggravated robbery in Limpopo Province: An analysis using multivariate techniques" prior to its final form. The data remained the property of the South African Police Service(SAPS).

3.14 Conclusion

The data and the methodology used during the research was enough to answer the research questions and the hypothesis(see Chapter 4). Above all, ethical procedures were followed during the execution of the research.

Chapter 4

Results and Discussions

4.1 Introduction

This chapter provides data analysis and discussions of the study of aggravated robbery in Limpopo Province in relation to poverty with the objective of answering the research questions. The data of aggravated crime used in the analysis was obtained from South African Police Service(SAPS). Aggravated robbery is defined as the use of force by a perpetrator to a victim in order to steal property usually causing bodily harm. Usually, the perpetrator will be armed. This crime propelled this research which focuses on another methodology of explaining the anatomy of aggravated robbery in Limpopo Province in relation to poverty. The method which was used is called Canonical Correlation Analysis(CCA). In this research, tests were employed to test for the statistical significance of the hypothesis. The variables were grouped into two sets, the criterion and predictor sets.

Poverty-aggravated robbery crime analysis presented below was run in SPSS(Version 22) software. The following discussion focused on two broadly areas, that is, an overview of aggravated robbery crime and interpretation of Canonical Correlation Analysis results.

4.2 Data analysis and Interpretation

4.2.1 Overview of aggravated robbery crime

Descriptive statistics gave a general insight of the variables under study. The analysis constitutes of offenders who were convicted of aggravated robbery crime in Limpopo Province from 1 January 2011 to 30 June 2013. The percentages of offenders as per district were as follows: Vhembe had 46.3%, Mopani had 23.6% and Capricorn had 30.1% (Table 4.2.1). The results showed a breakdown of aggravated robbery crime offenders for a sample of 559 people. Vhembe district was the highest among the three sampled districts.

Table 4.2.1: Aggravated offenders per district

District	Frequency	Per cent	Cumulative Per cent
Vembe	259	46.3	46.3
Mopani	132	23.6	69.9
Capricorn	168	30.1	100.0
Total	559	100.0	100.0

According to Figure 4.2.1, Vhembe district had high levels of common robbery and business robbery. Figure 4.2.1 showed Mopani district as having the least number of aggravated robbery crimes. This might be prompted by location of police stations which covered most of the rural communities with other types of criminal offences prevalent in those communities. Also it be attributed to the size of population or security strategies employed by SAPS in the district. For comparison's sake, Vhembe district topped up in the following categories such as common robbery, car hijacking and business robbery. It was evident that proportions of each criminal activity was not the same in each of the sampled districts. These disparities in proportion may be attributed to the geography of the districts (Shaw, 2002). Generally, most people in the province lives in the rural areas where some types of crime are not prevalent. Capricorn district may be affected by urbanisation (Shaw, 2002). Seshego and Polokwane clusters are located in Polokwane. Truck hijacking was more pronounced in Mopani District than in Vhembe and Capricorn districts. From the three districts,

house robbery, common robbery and business robbery were the most popular of offences committed by offenders. The three districts did not have offenders in bank robbery and cash-in-transit cases.

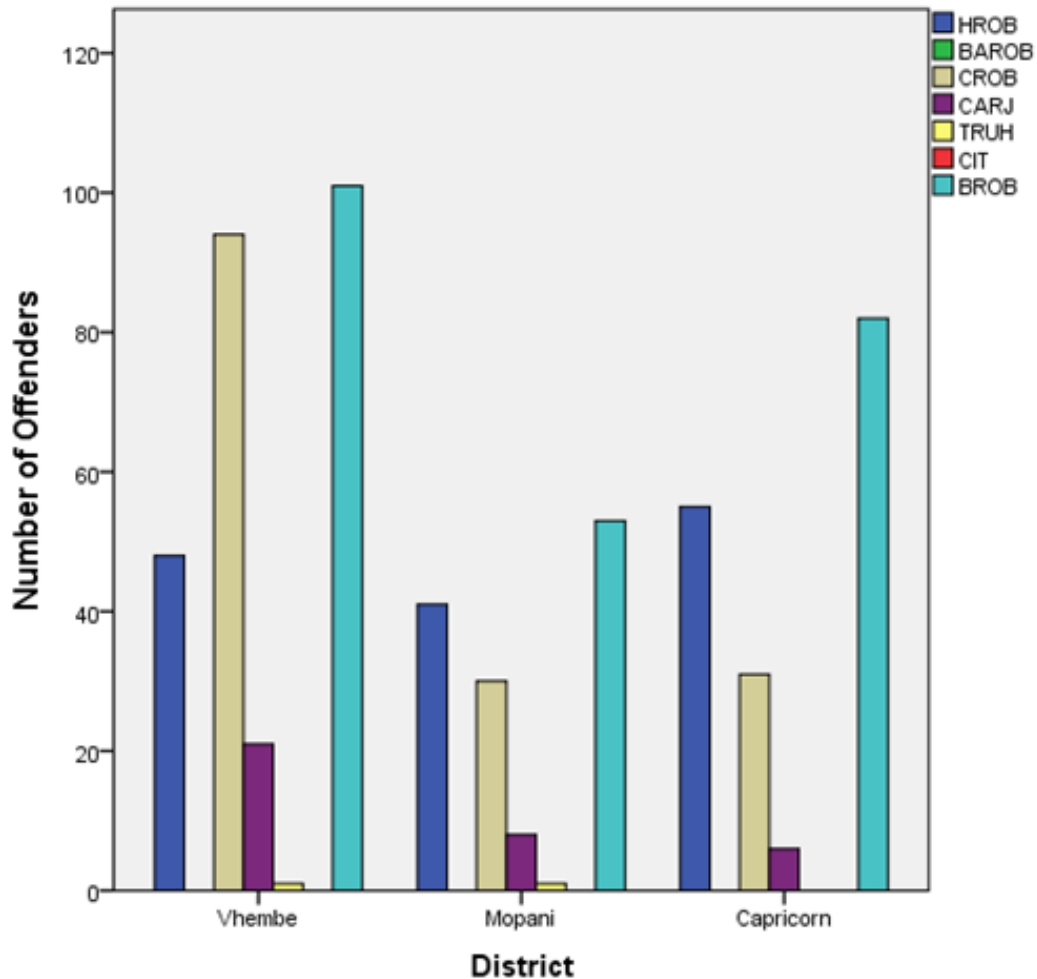


Figure 4.2.1: Comparison of types of aggravated crime per district

The pattern of aggravated crime per cluster also revealed that house robbery peaked in Thohoyandou, Polokwane and Tzaneen clusters. Musina had the lowest number of house robbery(Figure 4.2.2). From Figure 4.2.2, Thohoyandou, Makhado and Musina clusters had significant numbers of common robbery with Seshego recording the least number of common robbery crime. Then Seshego cluster which covers most of the Black townships in Polokwane had the highest number of business robbery crime

and was followed by Thohoyandou and Makhado. According to Figure 4.2.2, car hijacking was more pronounced in Thohoyandou cluster than in any of the clusters. Lebowakgomo didn't have offenders convicted of car hijacking. Also, the same figure showed Tzaneen and Thohoyandou clusters having a moderate number of offenders of truck hijacking. Cash-in-transit and bank robbery did not have offenders in any of these clusters. This might be due to a reduction of these crimes nationally. For example, in the past four years(2009/10 2012/13), cash-in-transit reduced by 62.4% and by 20.3% during the financial year 2012/13 at national level(SAPS, 2013). Also, bank robbery reduced by 93.1% for the past four years and by 80.0% during the past financial year of 2012/13(SAPS, 2013).

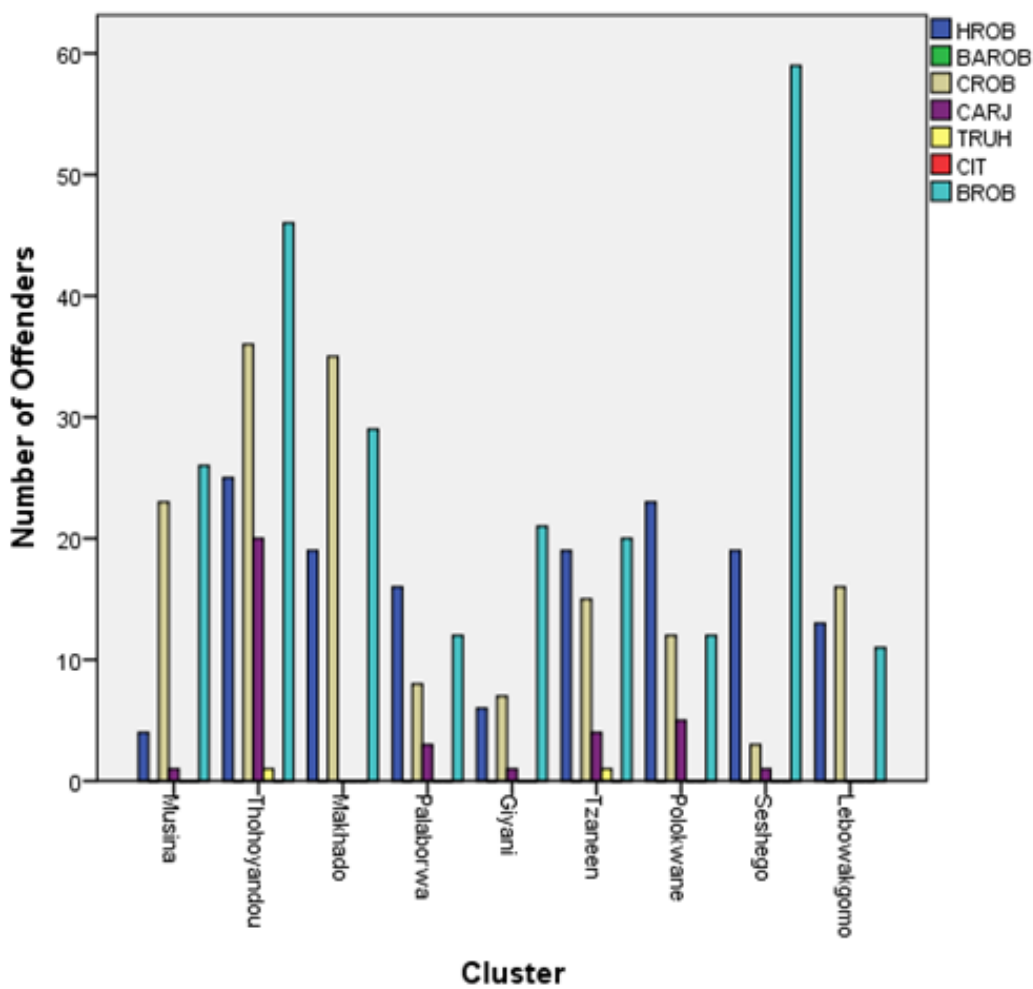


Figure 4.2.2: Aggravated crimes per cluster

From Table 4.2.2, 18(3.2%) females committed aggravated robbery crime as compared to males with a total of 541(96.8%) of this type of crime. The outcome does not contradict with the findings of Steffensneier and Allan(1996). They deduced from their research that male crime rates surpass female crime rates. It can be attributed to many factors such as cultural factors(Schmallegger, 2009). From the gender perspective, the power-control theory supports the notion of differences in crime rates(Bernard *et al.*, 2010).

Table 4.2.2 reveals that among the convicted offenders, 82.1% offenders were not employed. In addition, the employed offenders constituted 17.9% in this category. Unemployment as a structural factor might have propelled many to commit this crime(Jordan, 2004). This was not a surprise since unemployment in the province was pegged at 39.5% as indicated by the census of 2011(StatsSA, 2012). Table 4.2.2 also shows differences in proportion in terms of marital status. Weisburd and Waring(2001) affirmed that the probability of imprisonment is low for married people. The results in the research revealed this scenario of a low rate of married offenders. Furthermore, Table 4.2.2 shows a total of 116(20.8%) of married offenders and 443(79.2%) unmarried offenders. According to the national census of 2011, the percentages of married people, never married, widowed/widower and divorced/separated were 25.4%, 69.0%, 4.3% and 1.3% respectively(StatsSA, 2012). This explain the variability in criminal activities of offenders by marital status factor.

The disparity of race and crime as depicted by Table 4.2.2, shows a breakdown of convicts as 99.3% for Blacks and 0.7% for non-Blacks. This supports the notion that about 96.8% of the provincial population are blacks(StatsSA, 2012). Hence the probability of Black offenders is likely to be higher than that of other racial groups. Apartheid might be a contributing factor in terms of the inequality imbalances(Shaw, 2002; Greene and Gabbidon, 2012).

From the age perspective of poverty, poverty tends to increase as one grows older(Anyanwu,

2012). According to Anyanwu(2012), productivity has an inverse relationship with crime. As productivity decreases with age, crime increases as an individual attempts to have more money for the lost productivity(StatsSA, 2012). The figures in Table 4.2.2 also show that 36.9% of those who were convicted was 24 years or below. Those above 24 years were 63.1% which concurs with Anyanwu’s(2012) argument of inverse proportion of crime and age. The table also illustrates that a percentage of 45.1% attained primary education or they were illiterate. This was compared to 54.9% who had above primary education. A higher proportion of those who had above primary education might contradict with AEE(2006) where it was advocated that lower education propels high imprisonment. But this is not conclusive. Lochner and Moretti(2003) argued that an individual with high returns in criminal activities is most likely to spend more time in crime rather than work, irrespective of the level of education.

Table 4.2.2: Statistics of poverty levels of offenders

	Gender		Employment		Marital		RACE		Age		Education	
	Female	Male	Employed	Not Employed	Married	Not married	Non Black	Black	24 yrs or below	Above 24 yrs	Prim or illiterate	Above Prim
Total	18	541	100	459	116	443	4	555	206	353	252	307
Per cent	3.2	96.8	17.9	82.1	20.8	79.2	0.7	99.3	36.9	63.1	45.1	54.9

Statistics of aggravated robbery crime levels(house robbery, bank robbery, street/common robbery, carjacking robbery, truck hijacking, cash-in transit and business robbery) are shown in Table 4.2.3. House robbery crime was committed by 144(25.8%) offenders. Table 4.2.3 also shows statistics of offenders in each category of types of aggravated robbery as follows: 155(27.7%) committed common robbery(street/public robbery), 35(6.3%) committed carjacking robbery, 2(0.4%) committed truck hijacking

and 236(42.2%) committed business robbery. There were no offenders for bank and cash-in transit robberies.

Table 4.2.3: Types of aggravated crime committed by offenders

	House		Bank		Common		Carjacking		Truck		CIT		Business	
	Not committing	Committing	Not committing	Committing	Not committing	Committing	Not committing	Committing	Not committing	Committing	Not committing	Committing	Not committing	Committing
Total	415	144	559	0	404	155	524	35	557	2	559	0	323	236
Per cent	74.2	25.8	100	0	72.3	27.7	93.7	6.3	99.6	0.4	100	0	57.8	42.2

According to Figure 4.2.3, it was evident that the total number of aggravated robbery was highest for business robbery. It also depicts the three most common types of this crime. These were business robbery, common robbery and house robbery. Burger, Gould and Newham(2010) called them the “trio crimes”. There was also a notable difference between these crimes with bank and cash-in-transit which recorded zero.

The variation in the distribution of offenders among the types of aggravated robbery is due to apartheid-imposed geographies(Shaw, 2002). Apartheid ensured that Blacks couldn't go into White areas and policing was different between Black and White communities(Shaw, 2002). Shaw(2002) argued that city structures and crime patterns showed more crime activities in poorer townships(mostly for Blacks) and informal settlements. This didn't change much even in the post-apartheid era(Shaw, 2002). The culprits are Blacks. In this research it was evident that crime patterns were heavily depended on settlement structure, that is, between township areas, rural and informal settlements. For example, this was shown by types and number of offenders in Seshego, Lebowakgomo and Palaborwa clusters.

However, a lower record of bank robbery and cash-in-transit in Figure 4.2.3 may be attributable to effective security and arrest of prominent syndicates of these organised crimes. Even Shaw(2002) and Burger, Gould and Newham(2010), had the same view when they pointed out that arrests of well-known syndicates had an effect in crime reduction. On the other hand, few offenders of carjacking and truck hijacking may be attributed to the wide use of vehicle security technology and expertise of SAPS(Burger, Gould and Newham, 2010).

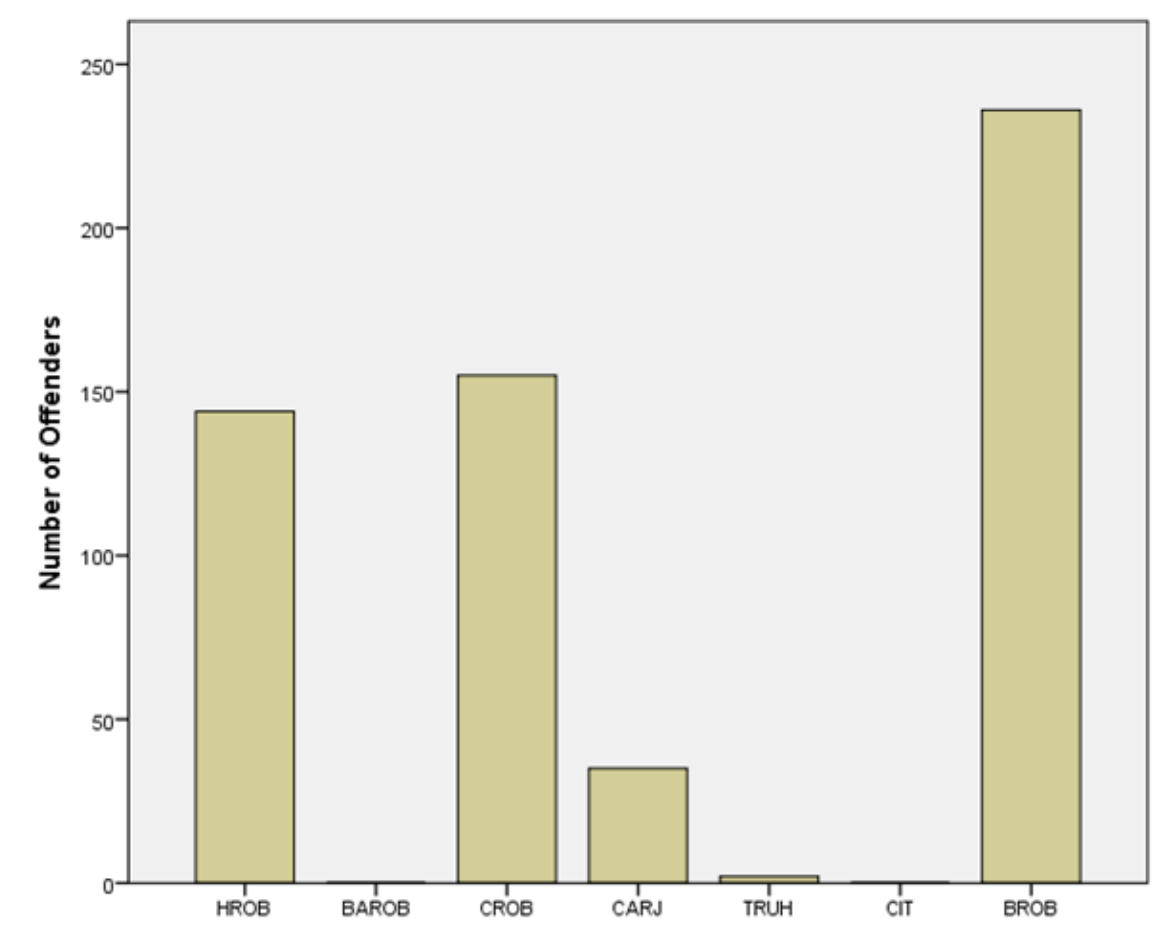


Figure 4.2.3: Graphical comparison of types of aggravated crime

From Table 4.2.4, Thohoyandou recorded the highest percentage(21.8%) of the total number of aggravated robbery. Clusters which recorded the lowest recordings of aggravated crime were Giyani (6.3%) and Lebowakgomo (6.8%) which was second from the bottom in the list. According to Table 4.2.4, Thohoyandou, Makhado and

Seshego showed that aggravated robbery crime is prevalent in these clusters. An equal number of offenders per cluster was targeted at first but this was derailed by unavailability of statistics in certain types of aggravated robbery. An in-depth study of the research variables is found in the next sections below.

Table 4.2.4: Cluster comparison of aggravated robbery offenders

Cluster	Frequency	Per cent
Musina	54	9.7
Thohoyandou	122	21.8
Makhado	83	14.8
Palaborwa	39	7.0
Giyani	35	6.3
Tzaneen	58	10.4
Polokwane	48	8.6
Seshego	82	14.7
Lebowakgomo	38	6.8
Total	559	100.0

4.2.2 Interpretation of Canonical Correlation Analysis results

Canonical Correlation Analysis was performed on poverty-aggravated robbery crime data with six predictor variables of poverty. The criterion set had five variables. Bearing in mind that the sensitivity of canonical correlation analysis to missing data might not produce desirable results, data screening was performed by leaving out the criterion variables, bank robbery and cash-in-transit. There were no convictions in these categories in the province during the given time frame(that is, two and half years) of the research.

However, CCA methodology does not require distributional assumptions(Malacarne,

2014). Hence, assumptions for hypothesis testing were not performed using the fact that data was assumed to have met requirements of multivariate normality and homogeneity of variance (Malacarne, 2014). In order to succeed in performing hypothesis testing in the context of the General Linear Models, data was transformed using the square root transformation. Normally, crime data follows a Poisson distribution. By so doing, the process transformed data into continuous multivariate data.

4.2.2.1 Testing the research hypothesis

Multivariate tests were performed to test whether the maximal correlation and those orthogonal successive correlations that follow were significantly different from zero. All the p values of the three tests, that is, the Pillai's criterion, Hotellings' trace and Wilks' test were less at .05 level of significance. The null hypothesis was rejected and we can conclude that canonical correlations were significantly different from zero. For example, $F(30, 2760.00)=1.563$, $p<.05$. Hence, this indicated that correlations were significantly different from zero. From this hypothesis test, the two sets of poverty and aggravated robbery crime were correlated. This answered the sub-question, "Is there any correlations between poverty and aggravated robbery crime criteria?" Table 4.2.5 portrays the results of the tests.

Table 4.2.5: Multivariate Tests of Significance

Test Name	Value	Approx. F	Hypoth. DF	Error DF	Sig. of F
Pillais	.08353	1.56314	30.00	2760.00	.026
Hotellings	.08644	1.57435	30.00	2732.00	.025
Wilks	.91854	1.57007	30.00	2194.00	.025
Roys	.04815				

4.2.2.2 Deriving canonical functions

Canonical variates of poverty and aggravated robbery crime were tested in hierarchal form. The research goal was to obtain canonical variates to interpret. The first pair of

canonical variate(functions 1 to 5) with a maximum canonical correlation of .219(Table 4.2.6 and Table 4.2.7) was extracted. It was statistically significant, $F(30.00, 2194.00) = 1.570$ and $p < 0.05$ (Table 4.2.6). Based on the residual variance, the other variates were tested in step-down order. Second(function 2 to 5), third(function 3 to 5), fourth(function 4 to 5) and fifth(function 5 to 5) canonical variates, had canonical correlations of .136, .100, .061 and .057 respectively. All these pairs of canonical covariates were not significant. From Table 4.2.6, the first pair of canonical covariate was obtained for further analysis. Even the magnitude of the canonical correlation of the first root pair, supported this notion. In the case of the magnitude of canonical correlations, Hair *et al.*, (2006), pointed out the absence of acceptable universal guidelines. Hence the discretion lies on the contribution of canonical correlations in answering the research problem (Hair *et al.*, 2006). Hence, this gives an answer to the main research question “Is there any meaningful relationship between poverty and aggravated robbery crime in Limpopo Province?” Given the magnitude of the problem of poverty and aggravated robbery crime in the province, it positively shows the existence of a meaningful relationship between poverty and aggravated robbery crime.

Table 4.2.6: Dimension Reduction Analysis

Roots	Wilks L.	F	Hypoth. DF	Error DF	Sig.of F
1 TO 5	.91854	1.57007	30.00	2194.00	.025
2 TO 5	.96500	.98363	20.00	1821.78	.479
3 TO 5	.98308	.78490	12.00	1455.45	.667
4 TO 5	.99303	.64361	6.00	1102.00	.695
5 TO 5	.99676	.89845	2.00	552.00	.408

4.2.2.3 Aggravated robbery crime canonical loadings analysis

Canonical loadings or structure coefficients were used in the analysis in order to assess the contribution of the original variables of aggravated robbery to the first canonical crime(aggravated robbery crime) variate. This implies that an assessment

Table 4.2.7: Eigenvalues and Canonical Correlations

Root No.	Eigenvalue	Pct.	Cum. Pct.	Canon Cor.	Sq. Cor
1	.05058	58.51506	58.51506	.21942	.04815
2	.01873	21.66794	80.18300	.13559	.01839
3	.01012	11.70896	91.89196	.10010	.01002
4	.00375	4.34211	96.23407	.06115	.00374
5	.00326	3.76593	100.00000	.05696	.00324

of correlations between individual variables with respect to their canonical variates is performed with a goal of extracting important variables. Conclusions were based on the magnitude of the loadings greater than the absolute value of 0.3(Kinnear and Gray, 2006). Table 4.2.8 showed that common robbery(SQRTCROB) with loading of -.915, was the primary contributor to the criterion variable set of aggravated robbery crime. Also the results revealed that business robbery(SQRTBROB) with loading of .759 was a secondary contributor to the same set. These variables were inversely related since their structure coefficients had different signs. The remaining variables had a minimal contribution to the to the crime canonical variate. Therefore, the canonical loading analysis sought to answer the sub-research question “In a set of aggravated robbery, which variable has the most and which has the least impact on creating meaningful relationship between poverty and aggravated robbery crime?”. However the research identified two significantly crimes, that is, common robbery and business robbery as the main drivers of aggravated robbery. This was slightly different from Burger, Gould and Newham(2010) who had an additional list of house robbery. According to the research it was discarded because it did not have a significant contribution. Results of canonical loadings are shown on Table 4.2.8. After transformation, variable coding changed. For example, the new code for HROB(house robbery) was SQRTHROB(\sqrt{HROB}).

Table 4.2.8: Canonical Loadings for aggravated robbery crime set

Variable	Function No.				
	1	2	3	4	5
SQRTHROB	-.03839	-.76927	.59712	.04048	.22035
SQRTCROB	-.91485	.29678	-.20131	.08311	-.16595
SQRTCARJ	.24427	.04303	-.39968	.70383	.53232
SQRTTRUH	-.07365	.39713	.31236	-.41351	.75386
SQRTBROB	.75859	.45166	.07459	-.24207	-.39546

4.2.2.4 Poverty canonical loadings analysis

Similarly, using the same argument, the predictor canonical loadings were used in the evaluation of the strength of the contribution to the predictor set of poverty. This sought to answer the sub-research question “In a set of variables of poverty, which variable has the most and which one has the least impact on creating a meaningful relationship between poverty and aggravated robbery crime?”. Then Table 4.2.9 showed the poverty canonical loadings which were used to answer this sub-research question. Structure coefficients of the poverty set, showed that gender(SQRTGENDER) with loading of .727 was the main contributor to the poverty set. Rural poverty is well defined in non-urban areas and most households are single-guardian(Jensen, 2009). The gender of the household breadwinner has a great bearing on the income of the household. Even in previous research findings showed gender as a best predictor of poverty which result in committing crime(Schmalleger, 2009). In addition, the contribution of marital status(SQRTMSTAT) with loading of .576 to the predictor set of poverty was not a surprise. The research results did not deviate from the previous empirical results which identified marital relationship as an influential key driver to poverty reduction(Anyanwu, 2012).

On the other hand, employment status(SQRTEMP) with loading of .337) was third in ranking in contributing the first canonical variate of poverty set. From the nature

of the structure coefficients, they were positively correlated with *Poverty*₁ variate. High unemployment and inequalities in incomes was seen as one of the drivers of poverty(Phogole, 2010). Therefore, the research revealed the key variables of the *Poverty*₁ variate.

Through the inspection of the loadings of the poverty set, it was revealed that there were other variables which did not contribute much to the predictor set. These were age, race and educational attainment with loadings of -.006, .164 and -.203 respectively. By using a threshold of loadings greater than 0.3, they were excluded in the list of important variables(Table 4.2.9).

Table 4.2.9: Canonical Loadings for the poverty set

Covariate	CAN. VAR.				
	1	2	3	4	5
SQRTGENDER	.72739	.30746	.07397	.10883	.15294
SQRTEMPS	.33747	-.47508	-.53076	.46145	.21178
SQRTMSTAT	.57617	-.43723	.39937	.14078	-.44925
SQRTRACE	.16424	.68802	-.30818	.17547	-.35478
SQRTAGE	-.00598	-.07306	-.59637	-.72788	-.28537
SQRTEdua	-.20311	-.15925	-.32269	.48725	-.61913

Figure 4.2.4 shows a pictorial diagram which displays associations of the two sets of variables. It is meant for easy visualisation of the poverty-aggravated robbery crime model. Different colours were used to differentiate the two sets, that is, poverty and aggravated robbery crime. The one direction arrows, showed causal correlation of variables with their respective canonical variates. For the double arrow representation, it depicts the reciprocal causal influences interrelationship between poverty and aggravated robbery.

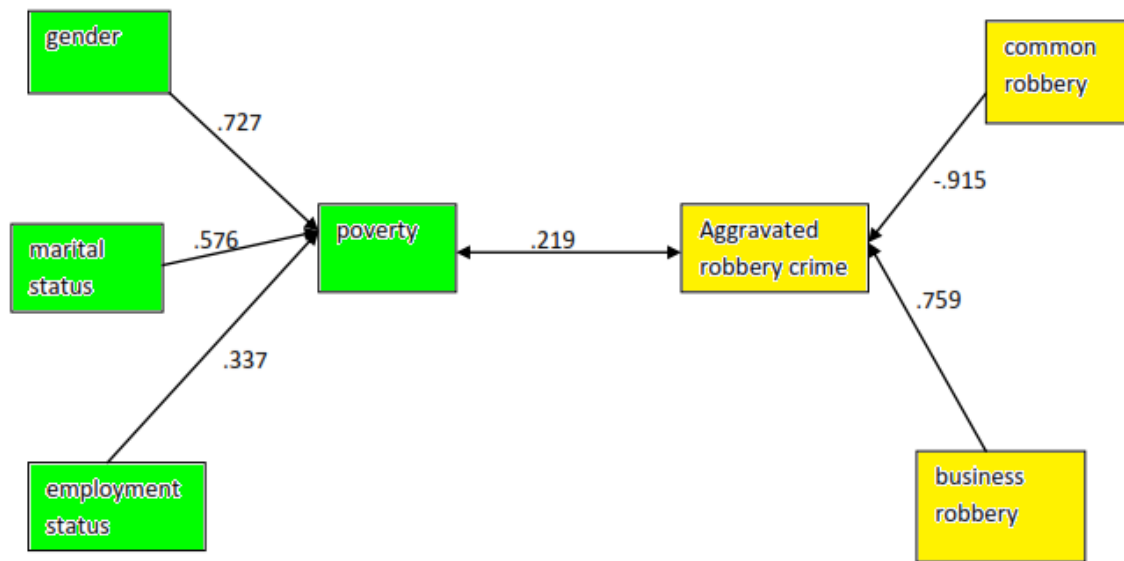


Figure 4.2.4: Canonical Correlation Analysis of Poverty and Aggravated robbery crime

4.2.2.5 Canonical cross-loadings of poverty-aggravated robbery crime datasets

An assessment was also made of the correlation of variables with the opposite canonical variate by the use of cross-loadings. Firstly, independent variables, that is, gender, marital status and employment status were correlated with aggravated robbery crime (with correlations of .160, .126 and .074 respectively). According to the research problem, magnitude of correlations were not strictly observed since the goal rests on the researcher in answering the research problem (Hair *et al.*, 2006). Table 4.2.10 shows the cross-loadings for the poverty set with the aggravate robbery crime set. In addition, these variables had direct relationship with aggravated robbery crime as depicted by signs of the cross-loadings.

In studying the cross-loadings of the aggravated robbery crime in relation to poverty, it can be seen that common robbery and business robbery exhibit correlation with poverty (-.201 and .166 respectively). Looking at the dependent's cross-loadings, they exhibits an inverse relationship. These cross-loadings are shown in Table 4.2.11.

Table 4.2.10: Cross-loadings for the poverty set

	1	2	3	4	5
SQRTGEND	.160	-.042	-.007	.007	.009
SQRTEMP	.074	.064	.053	.028	.012
SQRTMSTA	.126	.059	-.040	.009	-.026
SQRTRACE	.036	-.093	.031	.011	-.020
SQRTAGE	-.001	.010	.060	-.045	-.016
SQRTEUA	-.045	.022	.032	.030	-.035

Table 4.2.11: Cross-loadings for aggravated robbery crime set

	1	2	3	4	5
SQRTHROB	-.008	.104	-.060	.002	.013
SQRTCROB	-.201	-.040	.020	.005	-.009
SQRTCARJ	.054	-.006	.040	.043	.030
SQRTTRUH	-.016	-.054	-.031	-.025	.043
SQRTBROB	.166	-.061	-.007	-.015	-.023

4.2.2.6 Weights of the canonical variates

Interpretation of weights in connection to their respective canonical variates, was not considered in this research. This was due to the instability of weights (Hair *et al.*, 2006, Afifi, May and Clark, 2012). However, this section presented the canonical variates as discussed in the literature review. Canonical variates were expressed as

$$Crime_1 = -0.783*(SQRTCROB) + 0.307*(SQRTBROB) \quad (4.2.1)$$

and

$$Poverty_1 = 0.730*(SQRTGENDER) + 0.292*(SQRTEMP) + 0.502*(SQRTMSTAT) \quad (4.2.2)$$

The two tables (Table 4.2.12 and Table 4.2.13) shows standardised coefficients of the independent and dependent variables respectively.

Table 4.2.12: Standardized canonical coefficients for the poverty set

CAN. VAR.					
COVARIATE	1	2	3	4	5
SQRTGENDER	.72998	.30958	-.05940	.06806	.19879
SQRTEMP	.29227	-.40932	-.67581	.35594	.45673
SQRTMSTAT	.50168	-.40011	.46643	-.16032	-.63664
SQRTRACE	.15443	.73349	-.24604	.15263	-.35076
SQRTAGE	.23951	-.18594	-.53738	-.77960	-.29680
SQRTEDUA	-.28262	-.10782	-.19562	.52677	-.61008

Table 4.2.13: Standardized canonical coefficients for the aggravated robbery crime set

Function No.					
Variable	1	2	3	4	5
SQRTHROB	-.15339	.43120	2.23518	1.54061	-.32320
SQRTCROB	-.78252	1.37568	1.83780	1.70543	-.62715
SQRTCARJ	.16033	.64365	.64897	1.51415	.23825
SQRTTRUH	-.07937	.52191	.42712	-.29503	.70500
SQRTBROB	.30743	1.52430	2.16200	1.61851	-.78097

4.2.2.7 Redundancy analysis

For the independent set, the first canonical variate explains 17.4% of the variance in the poverty variables (see Table 4.2.14). It shows that the $Poverty_1$ variate was a fairly good predictor. Similarly, the dependent canonical variate shared an amount of variance of 29.6% with aggravated robbery crime variables (see Table 4.2.15). Hence, the $Crime_1$ variate was also a good predictor to its own set.

Table 4.2.14: Proportion of Variance of poverty set explained by its own Can. Var.

	Prop Var
CV-1	.174
CV-2	.169
CV-3	.167
CV-4	.174
CV-5	.143

Table 4.2.15: Proportion of Variance of aggravated robbery crime set explained by its own Can. Var.

	Prop Var
CV-1	.296
CV-2	.209
CV-3	.132
CV-4	.147
CV-5	.217

Another aspect of redundancy analysis, is to assess the amount of shared variance between the poverty canonical variate and the aggravated robbery crime variate. It is achieved by squaring the canonical correlation between the two sets (the independent set and dependent set). The first canonical variate pair accounts 4.80% amount of explained variance. With the SAPS zero tolerance to crime, this serves as an eye opener to crime reduction in relation to poverty-aggravated robbery's perspective.

Finally, the redundancy analysis focuses on the redundancy index analysis. In multivariate analysis, redundancy index analysis is used to confirm the practical significance of relationships and the success of canonical correlation analysis in uncovering the relationships (Hair *et al.*, 2006, Afifi, May and Clark, 2012). The proportion of variance in the poverty set explained by the canonical variate of aggravated robbery

crime was 0.8% (see Table 4.2.16). Looking at the aggravated robbery crime set of variables, the canonical variate of poverty explains 1.4%(see Table 4.2.17) of amount of variance in the dependent set. These low redundancy indexes(that is, 0.8% and 1.4%) is not a problematic issue. It is not a surprise since they were influenced by low canonical correlation and less amount of shared variance explained by their respective variates(Hair *et al.*, 2006). However, given the theoretical and practical perspective of the problem in the province relationships exhibited by the redundancy analysis are viewed as legitimate.

Table 4.2.16: Proportion of Variance of poverty set explained by aggravated robbery crime Can.Var.

	Prop Var
CV-1	.008
CV-2	.003
CV-3	.002
CV-4	.001
CV-5	.000

Table 4.2.17: Proportion of Variance of aggravated robbery crime explained by poverty Can. Var.

	Prop Var
CV-1	.014
CV-2	.004
CV-3	.001
CV-4	.001
CV-5	.001

4.3 Conclusion

In conclusion, this research exhibit a correlation of poverty and aggravated robbery in Limpopo Province with canonical correlation coefficient = 0.219. It was noticeable that gender, employment status and marital status were identified as the main components of poverty that triggered aggravated robbery crime in the province. This fulfilled the research objectives which sought to establish, and to quantify the correlation between poverty and aggravated robbery crime. Also an objective which sought to identify the main components of aggravated robbery crime was achieved. Common robbery and business robbery were identified as the main descriptors of aggravated crime. The canonical covariate of poverty yielded the theoretical expectations where unemployment for example, is viewed as a causative to poverty which then propel one to commit crime. This informative analysis, gives direction to variables which should be addressed with an objective of creating a free society from aggravated robbery and boosting economy in the right direction. The success of Canonical Correlation Analysis methodology in revealing interrelationships between poverty and aggravated robbery crime in Limpopo province, gives a signal to rightful implementation of intervention programmes which will curb poverty and aggravated robbery crime.

Chapter 5

Conclusions and Recommendations

5.1 Introduction

This chapter concludes with the key findings of the research and makes recommendations. The study focused on aggravated robbery offenders through docket analysis. The chapter reaffirms the research questions and seeks answers from the analysis done in previous chapters on results and discussions.

5.2 Conclusions

The key objective of the study was to explore the relationship between poverty and aggravated robbery crime levels by using one of the methods of multivariate analysis called Canonical Correlation Analysis. It was executed with the hope of injecting a significant input to the South African Department of Safety and Security in combating aggravated robbery crime in Limpopo Province. The study sought to answer the following research questions:

- Is there any meaningful relationship between poverty and aggravated robbery crime in Limpopo Province?

The sub questions which were used to answer the main research question above were:

- Is there any correlation between poverty criteria and aggravated robbery crime criteria?
- In a set of variables of aggravated robbery, which variable has the most and which

one has the least impact on creating a meaningful relationship between poverty and aggravated robbery crime?

- In a set of variables of poverty, which variable has the most and which one has the least impact on creating a meaningful relationship between poverty and aggravated robbery crime?

These questions were answered with aid of the following research objectives:

- (a) To examine the relationship between poverty levels and aggravated robbery crime in Limpopo province.
- (b) To determine whether the relationships between poverty levels and levels of aggravated robbery crime were significant.
- (c) To quantify the strength of the relationship of poverty and aggravated robbery crime.
- (d) To identify the main components of aggravated robbery crime and poverty that describes crime and poverty in the province.

5.2.1 Empirical Findings

This section synthesised the empirical findings in answering the research questions. The first two objectives, (a) and (b) were addressed and the empirical findings showed a significant correlation of poverty and aggravated robbery crime to be 0.219 with p -value of 0.025. This significant correlation was concluded after an examination of canonical correlation by using statistical tests such as the Hotellings' trace at 5 per cent significance. Hence, an answer to the main research question about the determination of the existence of a relationship of poverty and aggravated robbery crime was established. This provides an insight of the structure of the variables when it comes to short or long-term interventions.

Objective (c) cemented the conclusions about the main research question which were partly deduced from the first two objectives. The magnitude of the relationship between the two sets was examined by the use of redundancy index. The research findings depicted redundancy of 1.42% in the aggravated robbery set which was predicted by poverty variables. It was also concluded that aggravated robbery variables explained 0.8% of variance which was in the poverty set. Hence, it was concluded that the two sets, poverty and aggravated robbery had a meaningful association to which all stakeholders should not turn a blind eye.

However, another deduction was made from the empirical findings. Canonical correlation analysis addressed the last objective (d) which sought to determine the main and least significant components of aggravated robbery crime and poverty. Through an examination of canonical loadings (canonical structure correlations), the empirical findings showed common robbery as the most important contributor of aggravated robbery and the least being business robbery. On the other hand, gender, employment status and marital status of offenders provided a substantive contribution to the predictor set (poverty set). Empirical findings showed gender as the main component in the poverty set with employment status as the least important contributor. These empirical findings helped to answer the last of the sub questions of the research questions to which the research sought to find answers.

5.2.2 Theoretical implications

The research sought to contribute to the body of knowledge by finding variables of poverty that coalesce with aggravated robbery. The rationale of this research empirical model (poverty-aggravated robbery model) was embedded in the existing theories such as the General Strain Theory (GST). Gender, marital status and employment status were shown as triggers of aggravated robbery from the empirical findings. Broidy and Agnew (1997) also established the GST as the main driver of gender gap disparity in crime. Future aggravated robbery intervention programmes should focus on the promotion of achievement of positive valued goals on both males and females (an aspect of GST).

In previous empirical literature and even in this research, it has been revealed that marital status has an effect on crime (Visher *et al.*, 2009). This gives directions to the designing of programmes that encourage or improve partner relationships (Visher *et al.*, 2009). In the poverty-aggravated robbery model, employment status has a significant contribution in driving offenders to commit crime. The correlation of unemployment was also depicted in Melick's (2003) studies of motor vehicle theft and unemployment. These profound conclusions from empirical research would give an insight to the relevance of information recorded by SAPS and also serve as a pointer to crime prevention even before offending.

However, there was a slight difference in the main components of aggravated robbery crime. Burger, Gould and Newham (2010) identified house robbery, business robbery and common robbery as the key components of this crime. But in this research, only common robbery and business robbery were noted as the best descriptors of aggravated robbery. This difference might be attributed to different factors used to determine aggravated robbery.

From another perspective, to the academic society, Canonical Correlation Analysis has added more spices into multivariate analysis in analysing secondary data of offenders from dockets.

5.2.3 Policy implications

From the theoretical perspective of poverty and aggravated robbery, the empirical findings have given an insight to policy-makers of the relationship of these variables. The findings shed more light to policies, that is, the National Crime Prevention Strategy of 1996 and the 1998 White Paper on Safety and Security. Gender and marital status should be considered when implementing a policy of promoting economic development and growth in the informal sector. It can be noted that by avoiding the effects of the General Strain Theory, a significant drop in aggravated robbery might be noticed. Service delivery and eradication of poverty among the masses will lead

to a decrease in aggravated robbery. The poverty-aggravated robbery model plays a pivotal role in designing policies, for example, those which foster societal values like marriage. Since poverty breeds crime, from this research findings, aggravated robbery can be reduced by incorporating early positive developmental approaches(Emmett and Butchart, 2000).

5.3 Recommendations

The complexity of the research of the correlation of poverty and aggravated robbery crime is something which does not have closure to date. Hence, there is need for further research along this dimension. The following were noted:

- (i) Incomplete recording of the offender's information.

Some of the dockets did not have information such as employment status and level of education. In the research, offenders without full information were left out since canonical correlation is very sensitive to missing data. Hence, the SAPS Statistics Department are encouraged to fill in biographical information in dockets of all offenders.

- (ii) For future direction, it is suggested that research should focus on primary data captured from offenders. This will provide more inferences on the current quantitative research findings (empirical findings based on secondary data extracted from offenders dockets).

- (iii) The research also makes recommendations for an alternative methodology for improving the precision of relationships between the two sets of variables made by Canonical Correlation Analysis model. It suggested the Structural Equation Modeling(SEM) based on two principles. It uses a similar pictorial representations of associations and its ability to uncover complex patterns of causal relationships. SEM has an advantage over CCA since it can combine the latent(variable not directly measured) and observed variables. It has many advantages over most of the General Linear Models(GLM).

5.4 Concluding Remarks

In spite of the publicity of the impact of poverty in the province, poverty has enabled this researcher to make some inroads in finding solutions of the prevailing and persistent problem of aggravated robbery crime. The research revealed a significant relationship between poverty and aggravated crime. This will enable cross-breeding of ideas from all stakeholders in combating aggravated robbery crime by eradicating poverty in the province. As a matter of fact, poverty increases the vulnerability to crime.

Bibliography

- [1] AEE., 2006.Saving Future,Saving Future, Saving Dollars:The Impact of Education on Crime Reduction and Earnings.Washington D.C.Available at:<http://www.all4ed.org/files/SavingFutures.pdf>[Accessed on 26 May 2012]
- [2] Afifi A., May S and Clark V.A., 2012.Practical Multivariate Analysis.Fifth Edition.Boca Raton:Taylor and Francis Group, LLC:233-248
- [3] Anyanwu, J.C., 2012.Accounting for poverty in Africa:illustration with survey data from Nigeria.Available at:http://www.afdb.org/fileadmin/.../WPS%20Nos%20149_Anyanwu_Clean.pdf[Accessed on 20 July 2013.]
- [4] AudioEnglish.net Online Dictionary.“What does offender mean? definition, meaning and pronunciation...”,Available at:<http://www.audioenglish.net/dictionary/offender.htm>[Accessed on 22 December 2012]
- [5] Baker J.and Schuler N., (n.d).Analysis Urban Poverty:A summary of methods and approaches. Available at:<http://www.siteresources.worldbank.org/INTURBANPOVERTY/.../...>[Accessed on 2 October 2012]
- [6] Barker, L.,Pistang N. and Elliot R.,2002.Research Methods In Clinical Psychology:An Introduction for Students and Practitioners.Second Edition.Chichester:John Wley and Sons, Ltd:54-55.
- [7] Bernard T.J.,Snipes J.B. and Gerould A.L.,2010.VOLD’s Theoretical Criminology.Oxford:Oxford University Press.
- [8] Bhalla A.S.and Qui S., 2006.Poverty and Inequality Among Chinese Minorities.London:Routledge:1-15

- [9] Bhorat H. and Kanbur S.M.R., 2006. Poverty and policy in post-apartheid South Africa. Cape Town: Human Sciences Research Council: 288-294.
- [10] Bressler M.S., (n.d). The Impact of Crime on Business: A Model for Prevention, Decision and Remedy. *Journal of Management and Marketing Research*, 1-13. Available at: <http://www.aabri.com/manuscripts/09202.pdf> [Accessed on 27 May 2012]
- [11] Broidy L and Agnew R., 1997. Gender and crime: A General Strain Theory perspective. *Journal of research in crime and delinquency*. Vol. 34 No. 3, August. Sage Publications, Inc: 275-306 Available at: <http://troublesofyouth.pbworks.com/...Broidy+and+Agnew+-+Gender+and+Cr...> [Accessed on 12 January 2014]
- [12] Burchell J.M. and Milton J., 2005. Principles of Criminal Law, Third Edition. Lansdowne: Juta and Company Ltd: 40-41.
- [13] Burger J and Gould C., 2010. Violent crime in South Africa: What the latest crime statistics tell us. Available at: http://issafrica.org/crimehub/uploads/violentcrimeinsa_burger.pdf [Accessed on 14 January 2014]
- [14] Burger J., Gould C. and Newham G., 2009. The state of crime in South Africa: an analysis of the SAPS crime statistics for 2009/10. Available at: <http://www.issafrica.org/crimehub/uploads/BurgerGouldNewham.pdf> [Accessed on 19 November 2012]
- [15] Burger J., Gould C. and Newham G., 2010. CQ Number 25-Institute for security studies. Available at: <http://www.iss.co.za/uploads/CQ34Full.pdf> [Accessed on 11 November 2012]
- [16] Chow S., Shao J. and Wang H., 2003. Sample size calculations in clinical research. New York: Marcel Dekker.
- [17] Clark D., 1975. Understanding Canonical Correlation Analysis. London: Geo Abstracts Ltd.: 3-24

- [18] Cochran W.G., 1977.Sampling techniques(3rd ed).New York:John Wiley and Sons, Inc.
- [19] Cole G.F.,Smith C.E. and Jong C.D.,2012.The American System of Criminal Justice. Thirteenth Edition.Belmont:Wadsworth, Cengage Learning:1-43
- [20] Collins H., 2010.Creative Research:The theory and practice of research for the creative industries.Lausanne:AVA Publishing:120-121.
- [21] Creswell J.W., 2003.Research Design:Qualitative,Quantitative,and Mixed Methods Approaches.Second Edition.California:Sage Publications, Inc:3-162
- [22] CSV.R.,2010.Available at:<http://www.csvr.org.za/docs/study/CSV.Rstatement>[Accessed on 21 June 2012]
- [23] Detotto C.and Otranto E., 2010.“Does Crime Affect Economic Growth?” *KYKLOS*, Vol 63-August 2010 No 3,330-345.Available at: <http://www.onlinelibrary.wiley.com>...>Kyklos>Vol63Issue3>[Accessed on 29 July 2012]
- [24] Ebenezer O.and Iyaniwura J.O.,2012.“Canonical Correlation Analysis of Poverty and Literacy in Ekiti State, Nigeria”.*Mathematical Theory Modeling*. Vol.2,No.6.2012:31-38.Available at: <http://www.iiste.org/Journals/index.php/MTM/article/download/.../2070>[Accessed on 18 November 2012]
- [25] Emmett T.and Butchart A., 2000.Behind the mask:Getting to Grips With Crime and Violence in South Africa.Pretoria:HSRC Publishers Pretoria:3-21.
- [26] Gabbidon S.L., 2010.Criminological Perspectives on Race and Crime, Second Edition.New York:Taylor and Francis:1-33.
- [27] Geis J.A., 2012.Adaptive Canonical Correlation Analysis with considerations for high dimensional matrices:A weighted rank selection criterion approach with a HIV/Neurocognitive approach.Available at:<http://stat.fsu.edu/dissertation/Gwis.pdf>[Accessed on 13 December 2014]

- [28] GLM., 2012.Generalized Linear Models.Available at:<http://userwww.sfsu.edu/efc/classes/.../Generalized%20Linear%20Models.htm>[Accessed on 11 December 2014]
- [29] Golafshani N., 2003. “Understanding Reliability and Validity in Quantitative Research”.*The Qualitative Report. Volume 8 Number 4 December 2003*:597-607.Available at: <http://peoplelearn.homestead.com/medhome/qualitative/reliab.validity.pdf>[Accessed on 5 November 2012]
- [30] Govender P.,Kambaran N.,Patchett N.,Ruddle A.,Torr G.and Zyl N.V.,(n.d).Poverty and Inequality in South Africa and The World.Available at: <http://www.assa.ukubona.co.za/.../8d1f7275-bcb8-40e9-85d0-aebb9bdea3c6.pdf>[Accessed on 30 September 2012.]
- [31] Greene H.T and Gabbidon S.L., 2012.Race and Crime.Third edition.London:SAGE Publications, Inc.
- [32] Hagan F.E., 2011.Introduction to Criminology: Theories, methods, and criminal behaviour.California:SAGE Publication, Inc:375-377.
- [33] Hair(Jr) J.F., Black W.C., Babin B.J.and Anderson R.E., 2006.Multivariate Data Analysis.New Jersey:Pearson Prentice Hall.
- [34] Handley C.,(n.d).“Validity and Reliability in Research”.Available at:http://www.natco1.org/research//Validity-ReliabilityResearchArticle_000...[Accessed on 9 November 2012]
- [35] Haughton J and Khandker S.R., 2009.Handbook on Poverty and Inequality.Washington, DC:The World Bank.
- [36] Henry G.T., 1990. Practical Sampling.Newbury Park.Sage Publications,Inc.
- [37] Ho S., 2010.Explanatory Factor Analysis and Canonical Correlation.Trinity Western University.Available at:http://twu.seanho.com/10fall/cpsy501/.../12-EFI_CanonCorr.pdf[Accessed on 24 November 2012]

- [38] Israel G.D., 1992. Determining Sample *Size*¹. Fact Sheet PEOD-6, University of Florida. Available at: <http://edis.ifas.ufl.edu/pd006> [Accessed on 15 February 2013]
- [39] Jeng S.K., 2006. Canonical Correlation Analysis. Available at: <http://cc.ee.ntu.edu.tw/~skjeng/MultivariateAnalysis2006/CanonicalCorrelation..> [Accessed on 15 December 2014]
- [40] Jensen E., 2009. Teaching with poverty in mind: What being poor does to kids' brains and what schools can do about it. Alexandria: ASCD 6-7.
- [41] Jacobs E., Punt C., Ramathoka M. and Masekoameng A., 2009. A profile of the Limpopo Province: Demographics, Poverty, Income, Inequality and Unemployment from 2000 till 2007. PROVIDE Project Paper. Available at: http://www.elsenburg.com/provide/.../background/BP2009_1_9_%20LP%20 [Accessed on 27 August 2013]
- [42] Jordan G., 2004. The Causes of Poverty-Cultural vs. Structural: Can There Be a Synthesis? Available at: <http://www.asu.edu/mpa/Jordan.pdf> [Accessed on 11 July 2013.]
- [43] Kinnear P.R and Gray C.D., 2006. SPSS 14 Made Simple. New York: Psychology Press.
- [44] Kok P., 1998. South Africa's Magnifying Glass: A Profile of Gauteng. Pretoria: Human Science Research Council: 237-250.
- [45] Kothari C.R., 2004. Research Methodology: methods and techniques. Second Edition. New Delhi: New Age International (P) Ltd: 31-66.
- [46] Krejcie R.V. and Morgan D.W., 1970. Determining sample size for research activities. Available at: <http://opa.uprrp.edu/InvInsDocs/KrejcieandMorgan.pdf> [Accessed on 26 December 2012]
- [47] Kyei K.A. and Gyekye K.B., 2011. "Determinants of Unemployment in Limpopo Province in South Africa: Exploratory Studies". *Journal of Emerging Trends in Economics and Management Sciences (JETEMS)* 2(1): 54-61. Available at: <http://>

- [//jetems.scholarlinkresearch.org/.../...](http://jetems.scholarlinkresearch.org/.../...)[Accessed on 27 November 2012]
- [48] Levine M.S., 1977. Canonical Analysis and Factor Comparison. California: SAGE Publications, Inc:1-35.
- [49] Lochner L. and Moretti E., 2003. The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. Available at: <http://www.emlab.berkeley.edu/~moretti/Im46.pdf> [Accessed on 26 May 2012]
- [50] Malacarne R.L., 2014. Canonical Correlation Analysis. *The Mathematica Journal*. Wolfram Media, Inc. Available at: <http://www.mathematica-journal.com/data/uploads/2014/06/Malacarne.pdf> [Accessed on 27 November 2014]
- [51] Manly B.F.J., 1986. Multivariate Statistical Methods: A Primer. Chapman and Hall London.:114-125.
- [52] Masika., de Haan A. and Baden S., 1997. "Urbanisation and Urban Poverty: A Gender Analysis". *BRIDGE Report No. 54*. Brighton: Institute of Development Studies. Available at: <http://www.bridge.ids.ac.uk/reports/re54.pdf> [Accessed on 2 October 2012]
- [53] Melick, M.D., 2003. The relationship between crime and unemployment. *The Park Place Economist*. Illinois Wesleyan University. Vol. 11. Available at: <http://digitalcommons.iwu.edu/igi/viewcontent.igi?article=1069&context...> [Accessed on 5 January 2014]
- [54] Nykiel R., 2007. Handbook of Marketing Research Methodologies for Hospitality and Tourism. Binghamton: The Haworth Press. Inc:55-88.
- [55] ODC., 2002. Country Profile on Drugs and Crime. Available at: http://www.unodc.org/pdf/southafrica/country_profile_southafrica.pdf [Accessed on 2 June 2012]
- [56] Penslar R.L., 1995. Research Ethics: Cases and Materials. Bloomington: Indiana University Press:13-23.

- [57] Phogole M.,2012.Issues of Increasing Levels of Poverty and Hunger in Africa with specific reference to South Africa.Available at:<http://www.ai.org.za/wp-content/plugins/download.../download.php?id=22>[Accessed on 3 October 2012]
- [58] Punt C.,Pauw K.,Schoor M.,Nyhodo B.,McDonald S.,Chant L.and Valente C.,2005.A profile of the Limpopo province: Demographics, poverty, inequality and unemployment.*Elseburg*.Available at: http://www.elseburg.com/.../BP2005_1_9%20Demographics%20LP.pdf[Accessed on 2 November 2012]
- [59] Rauch J.(n.d).Changing step:crime prevention policy in South Africa.Available at:<http://www.issafrica.org/CJM/corrections/prisoncrowding/.../rauchbook.pdf>[Accessed on 31 May 2012]
- [60] Rencher A.C., 2002.Methods of Multivariate Analysis.Second Edition.John Wiley and Sons, Inc: 5-379.Available at: <http://tx.shu.edu.tw/.../!DataAnalysis/Methods%20of%20Multivariate%20Anal...>[Accessed 15 February 2014]
- [61] Ricketson(Sr) R.S., 2008.An exploration of the relation of leadership styles and dimensions of courageous followership.Ann Arbor:ProQuest LLC.:71-72.
- [62] Ross J.I.,2012.An Introduction to political crime.Bristol:The Policy Press:ix-8.
- [63] Sampson R.J.,Laub J.H. and Wimer C., 2006.Does Marriage Reduce Crime?A Counterfactual Approach To Within-Effects.*Criminology*.Volume 44. Number 3.Available at:http://www.scholar.harvard.edu/sampson/files/2006_criminology_laubwimer.pdf[Accessed on 26 May 2012]
- [64] SAPS., 2010.South African Police Service.Available at:<http://www.info.gov.za/view/DownloadFileAction?id=150105>[Accessed on 29 May 2012]
- [65] SAPS., 2013.Crime statistics overview RSA- 2012/2013.Available at:http://www.saps.gov.za/statistics/.../crime_statistics_presentation.pdf[Accessed on 31 December 2013]

- [66] SARPN), (n.d). Section 3: The causes of poverty and addressing the challenges. Available at: http://www.sarpn.org/CountryPovertyPapers/.../taylor/report4_section3.pdf [Accessed on 3 October 2012]
- [67] Schmallegger F., 2009. *Criminology Today: An Integrative Introduction*, Fifth Edition. New Jersey: Person Prentice Hall: 6-507.
- [68] Sekaran U. and Bougie R., 2010. *Research Methods for Business: A skill building approach*. Fifth Edition. New Delhi: Wiley India Pvt. Limited: 69-73.
- [69] Shaw M., 2002. *Crime and policing in post-apartheid South Africa: Transforming under fire*. London: Hurst and Co. (Publishers) Ltd: 50-51
- [70] Sherry A and Henson R.K., 2005. Conducting and Interpreting Canonical Correlation Analysis in Personality Research: A User-Friendly Primer. *Journal of Personality Assessment*, 84(1), 37-48. Available at: <http://digitalparadigm.net/Staging/alissa/download/PDFs/CCA.pdf> [Accessed on 14 February 2013]
- [71] Siegel L.J., 2010. *Introduction to Criminal Justice*, Twelfth Edition. Boston: Wadsworth, Cengage Learning: 71-74.
- [72] Singleton R.A., Straits B.C. and Straits M.M., 1993. *Approaches to Social Research*, Second Edition. Oxford: Oxford University Press, Inc: 451-494.
- [73] StatsSA., 2012. *Census 2011 Municipal report Limpopo: Statistics South Africa*. Available at: http://www.statssa.gov.za/Census2011/Products/LP_Municipal_Report.pdf [Accessed on 9 September 2013]
- [74] SPII., 2007. *The measurement of poverty in South Africa Project: key issues*. Available at: <http://www.treasury.gov.za/publications/.../povertyline/SPII%20document...> [Accessed on 1 October 2012]
- [75] Strategic Policy Brief., 2009. *Theories of the causes of crime*. Ministry of Justice. Available at: <http://www.justice.gov.nz/justice...crime/.../spb-theories-on-the-causes-of-crime> [Accessed on 28 November 2013]

- [76] Sukamolson S.,(n.d).Fundamentals of quantitative research. Available at:<http://www.culi.chula.ac.th/e-journal/bod/suphat%20sukamolson.pdf>[Accessed on 14 October 2012]
- [77] Thabane L; Ma J; Chu R; Cheng J; Ismaila A; Rios L.P; Robson R; Thabane M; Giangregorio L and Goldsmith C.H.,(2010).A tutorial on pilot studies: the what, why and how.BMC Medical Research Methodology.Available at:<http://www.biomedcentral.com/content/pdf/1471-2288-10-1.pdf>[Accessed on 27 July 2013]
- [78] Tshiwula L., 1998.Crime and delinquency.Pretoria:Kagiso Publishers:1-22.
- [79] UM.,(n.d).Canonical Correlation Analysis.Available at:<http://www.education.umd.edu/EDMS/.../CanonicalCorrelationChapter.pdf>[Accessed on 30 November 2012]
- [80] University of Pretoria.,2003.Secondary data analysis.Available at:<http://web.up.ac.za/sitefiles/file/43/Circular09-03.pdf>[Accessed on 28 November 2012]
- [81] Urdan T.C.,2005.Statistics in plain English.New Jersey:Lawrence Erlbaum Associate, Inc:117-118.
- [82] Visher C.A; Knight C.R; Chalfin A and Roman J.K., 2009.The Impact of Marital and Relationship Status on Social Outcomes for Returning Prisons.Available at:http://www.urban.org/UploadedPDF/411871_returning_prisoners.pdf [Accessed on 27 December 2013]
- [83] WBI.,2005.Introduction to poverty analysis.*Poverty Manual, All, JA Revision*.Available at:<http://siteresources.worldbank.org/PGLP/Resources/PovertyManual.pdf>[Accessed on 15 January 2015].
- [84] Weatherburn D.,2001.What Causes Crime?*Crime and Justice Bulletin, Number 54*:1-12.Available at:http://www.cso.nsw.gov.au/lawlink/bocsar/II_bocsar.nsf/.../cjb54.pdf[Accessed on 21 May 2012]

- [85] Weijun T., 2008. Research Methods for Business students Chapter 7. Available at: http://site.iugaza.ed.ps/walhabil/files/2010/02/Chapter_7.pdf [Accessed on 23 November 2012]
- [86] Weisburd D. and Waring E.J., 2001. White-Collar Crime and Criminal Careers. Cambridge: Cambridge University: 116-117.
- [87] Wimmer R.D. and Dominick J.R., 2011. Mass Media Research: An Introduction, Ninth Edition. Boston: Wadsworth, Cengage Learning: 44-45.
- [88] Yount R., 2006. Factorial and Multivariate Analysis of Variance. Available at: http://www.napce.org/documents/research-design.../25_factorial_4th.pdf [Accessed on 10 June 2012]

Appendix A

Supervisor's letter

Prof John. O. Olaomi
Department of Statistics
TvW Building, Room 7 -115
Pretoria

12 October, 2012

The Provincial Commissioner
South African Police Service (SAPS)
Limpopo Province

Dear Sir/Ma

This letter serves to confirm that Mr T. RWIZI, Student Number 37348132 is a registered student at UNISA and is studying towards the Masters Degree in Statistics.

Mr RWIZI thesis is on the “**Canonical Correlation Analysis of Aggravated Crime and Poverty in Limpopo Province**” and he needs to collect data on aggravated crimes in Limpopo Province. I shall therefore be grateful, if you could give him free access to the relevant data.

Should you have any further queries, please feel free to contact me at telephone number 012 429 3936 or email: olaomjo@unisa.ac.za

Thank you.



Prof JO Olaomi
Associate Professor of Statistics

Appendix B

SAPS Provincial and National letters



Verwysing Reference	2/1/2/1
Navrae Enquiries	Col Mashile Major Montjane
Telefoon Telephone	015 290 6090/ 6202
Faksnomme r Fax number	015 230 1023


ORGANIZATIONAL DEVELOPMENT
SOUTH AFRICAN POLICE
POLOKWANE
LIMPOPO

A: Mr T Rwizi
Tshivhase Secondary Shool
Private Bag x1168
Vhufuli
0971

B: National Head: (ATT: MAJOR GENERAL MENZIWA)
Strategic Management
South African Police Services
Pretoria
0001

**AUTHORITY TO CONDUCT RESEARCH: ANALYSIS USING MULTIVARIATE
TECHNIQUES IN LIMPOPO: MASTERS DEGREE STUDY: UNISA: RESEARCHER
MR T RWIZI**

- A.1. Your application for authority to conduct the research indicated above refers.
2. The authority is granted with conditions that such research is conducted within the given scope.
3. You are advised to adhere to the provisions of the National Instruction 1/2006 on Research in the SAPS.
- B.1 Copy for your information.


MAJOR GENERAL
DÉPUTY PROVINCIAL COMMISSIONER
OPERATIONS OFFICER
LIMPOPO PROVINCE
L.J MASHILO

DATE: 2012-11-26

SUID-AFRIKAANSE POLISIEDIENS



SOUTH AFRICAN POLICE SERVICE

Privaatsak/Private Bag X94

Reference Nr Verwysing	3/34/2
Navrae Enquiries	Col J Schneller Lt Col GJ Joubert
Telefoon Telephone	012-393 3177 012 393 3118
Faksnommer Fax number	012-393 3178

STRATEGIC MANAGEMENT COMPONENT
HEAD OFFICE
PRETORIA


The Provincial Commissioner
Limpopo

(Attention Col Mashile)

RE: RESEARCH REQUEST: AN ANALYSIS USING MULTIVARIATE TECHNIQUES IN
LIMPOPO: MASTERS DEGREE STUDY: UNISA: RESEARCHER: MR T RWIZI

- 1 The research proposal of Mr T Rwizi pertaining to the above-mentioned topic refers
- 2 The aim of the research is to determine a mathematical forecasting multiple regression model for aggravated robbery crime in Limpopo
- 3 The researcher will use a multiple regression analysis to formulate prediction models. Data will be collected for truck hijacking, house robbery, and business robbery, cash-in-transit (CIT) robbery, carjacking, bank robbery, street/public robbery. Data for these crimes will be collected according to the following variables: age, marital status, educational attainment, employment status, race and genders
- 4 This office recommends the research in terms of National Instruction 1 of 2006 Research in the Service, subject to the approval of the Provincial Commissioner Limpopo

With kind regards


MAJOR GENERAL
HEAD: STRATEGIC MANAGEMENT
M MENZIWA

Date: *2012/04/20*

Appendix C

Data coding

Aggravated robbery crime research data was excluded from the appendix due to space. However, for a clear understanding of how it was recorded the following coding was used. For aggravated robbery variables, 0 was coded for not committing crime and 1 for committing crime. Coding for poverty variables and districts is shown Table C.0.1 and Table C.0.2 respectively.

Table C.0.1: Coding format for poverty variables

Variable	Coding
Gender	0 if the offender is female
	1 if the offender is male
Employment status	0 if employed
	1 if not employed
Marital status	0 if married
	1 if not married
Race	0 if Non Black
	1 if Black
Age	0 if age 24 years or below
	1 if age is above 24 years.
Educational attainment	0 if educational attainment is above Primary education
	1 if educational attainment is Primary education or illiterate

Table C.0.2: Districts and clusters coding

District	District code	Name of Cluster	Cluster code
Vhembe	1	Musina	1
		Thohoyandou	2
		Makhado	3
Mopani	2	Palaborwa	4
		Giyani	5
		Tzaneen	6
Capricorn	3	Polokwane	7
		Seshego	8
		Lebowakgomo	9