

AN ECOLOGICAL ADJUSTED RANDOM EFFECT MODEL FOR VIOLENT AND
PROPERTY CRIME IN WINDHOEK (2011-2016)

A RESEARCH THESIS SUBMITTED IN PARTIAL
FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN APPLIED STATISTICS AND DEMOGRAPHY

OF

THE UNIVERSITY OF NAMIBIA

BY

JONAS AMUNYELA

(200745077)

APRIL 2020

MAIN SUPERVISOR: DR. I. NEEMA (NAMIBIA STATISTICS AGENCY(NSA))

CO-SUPERVISOR: PROF. L. PAZVAKAWAMBWA (UNAM)

Abstract

Count data that are zero inflated are often analysed using Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) when observations are correlated in ways that require random effects. The present study investigates ecological factors influencing Violent and Property crime in Windhoek by using data obtained from the Windhoek police over the period of six consecutive years (2011 to 2016). The ecological concepts were measured at several different levels of aggregation. Limited studies in Windhoek have considered analysing crime data on a newly established Generalized Linear Mixed Model via Template Model Builder (TMB) R-package. The researcher considered the number of reported Property and Violent crime for the study period as a quantitative design. Crime was counted with respect to Month, Season, Year, Location and Density. Through an exploratory study, it was found that both Property and Violent crime data contained more zeros than would be expected. Furthermore, in specifying the probability distribution using confidence interval, the researcher found out that the Negative Binomial distribution was appropriate for the two types of crime. Besides that, the lognormal distribution also appears to be an appropriate distribution for modeling Violent crime. However, when comparing models fitted in the context of these two distributions it was found that the Relative Risk (RR) were highly significant for models fitted via Negative Binomial distribution. By adopting a ZINB-GLMM, the study attempts to address the potential covariates for both Property and Violent crime. The study shows that most of the variation in the study of Property and Violent crime was due to locations. On average more Property crime (68%) was committed in Windhoek as compared to Violent crime (32%). Crime was high during Spring and Winter time during the study period. The study further discovered that areas with high population densities have a high crime intensity.

Key words: Crime intensity; Random effects; Windhoek; Zero-inflated; ZINB-GLMM

Table of Contents

Abstract	i
List of Tables	iv
List of Figures	v
LIST OF ABBREVIATIONS AND/ACRONYMS	vii
ACKNOWLEDGEMENTS	ix
DEDICATIONS	x
DECLARATIONS	xi
CHAPTER ONE	1
1. Introduction	1
1.1 Background of the study	1
1.2 Statement of the problem.....	5
1.3 Objectives of the study	5
1.4 Significance of the Study.....	6
1.5 Limitation of the study	6
CHAPTER TWO	7
2 Literature Review	7
2.1 Introduction	7
2.2 Consequence of Violent and Property crimes	8
2.3 Factors influencing Violent and Property crimes	8
2.4 Statistical modelling of Violent and Property crimes.....	9
2.4.1 Zero Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM)	12
2.4.2 Poisson model.....	16
2.4.3 Negative Binomial (NB) model.....	17
2.4.4 Zero-Inflated Negative Binomial (ZINB) model.....	17
2.5 Model selection.....	19
2.5.1 Akaike Information Criteria (AIC).....	19
2.5.2 Residual plot for ZINB-GLMM	19
CHAPTER THREE	21
3. Research Methodology	21
3.1 Research Design	21
3.2 Population.....	21
3.3 Sample	21
3.4 Research instrument	21

3.6.	Research Ethics.....	24
CHAPTER FOUR.....		25
4.	DATA ANALYSIS AND PRESENTATION OF THE RESULTS.....	25
4.1	Introduction	25
4.3	Specifying the suitable model that best fits the data	37
4.3.1	Specifying a suitable model for the Number of Property crime.....	37
4.3.2	Specifying a suitable model for the Number of Violent crimes	40
4.5.1	Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) for the Number Property crime.....	44
4.5.2	Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) for the Number Violent crime.....	49
CHAPTER FIVE		53
5.	DISCUSSIONS, CONCLUSIONS AND RECOMMEDATIONS	53
5.1	DISCUSSIONS.....	53
5.2	CONCLUSION.....	55
5.3	RECOMMENDATIONS.....	56
References.....		58
Appendix A.....		64
Appendix B.....		73

List of Tables

TABLE

Table 3-1: Summary of crime statistics across seasons and years.....	23
Table 4-1: Comparing AIC for five Property and Violent crime models	42
Table 4-2a: Summary of the ZINB-GLMM model for Property crime.....	46
Table 4-2b: Summary of the ZINB-GLMM model for Violent crime.....	50

List of Figures

Figure 4.1a: Histogram for the Number of Violent crimes.....	27
Figure 4.1b: Histogram for the Number of Property crimes.....	27
Figure 4.2a: A box plot for the Number of Violent crimes	28
Figure 4.2b: A box plot for the Number of Property crimes	28
Figure 4.3a: Q-Q plot for the Number of Property crime.....	29
Figure 4.3b: Q-Q plot for the Number of Violent crime.....	30
Figure 4.4a: Number of Violent crime distribution for each year	31
Figure 4.4b: Number Property crime distribution for each year.....	31
Figure 4.5: The Number of Property crime distribution for each month (January to December).....	33
Figure 4.6a: Number of Property crimes for each season	34
Figure 4.6b: Number of Violent crimes for each season	34
Figure 4.7a: Plotting Number Property crimes against locations.....	35
Figure 4.7b: Plotting Number Violent crimes against locations.....	36
Figure 4.8a: Assessing if the Number of Property crime data follows a normal distribution using confidence interval	37
Figure 4.8b: Assessing if the Number of Violent crime data follows a normal distribution using confidence interval.....	40

Figure 4.9a: Assessing if the Number of Property crime data follows a log-normal distribution using confidence interval.....38

Figure 4.9b: Assessing if the Number of Violent crime data follows a log-normal distribution using confidence interval.....41

Figure 4.10a: Assessing if the Number of Property crime data follows a negative binomial distribution using confidence interval.....39

Figure 4.10b: Assessing if the Number of Violent crime data follows a negative binomial distribution using confidence interval.....41

Figure 4.11a: Residual plot for Number of Property crimes model (ZINB-GLMM)43

Figure 4.11b: Residual plot for Number of Violent crimes model (ZINB-GLMM)44

LIST OF ABBREVIATIONS AND/ACRONYMS

ATM	Automatic Teller Machine
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CoW	City of Windhoek
CSV	Comma-Separated Values
GLMM	Generalized Linear Mixed Model
glmmTMB	Generalized linear mixed model Template Model Builder
INLA	Integrated Nested Laplace Appproximation
ML	Maximum Likelihood
NUMBEO	Collaborative online database
NSZ	Non-Structural Zero
OLS	Ordinary Least Square
NB	Negative Binomial
Q-Q	Quantile-Quantile plot
RML	Restricted Maximum Likelihood
SPSS	Statistical Package for the Social Science
UN	United Nations
UNAM	University of Namibia
UNODC	United Nations Office on Drugs and Crime

REC Research Ethics Committee

WCPOLS Windhoek City Police

ZINB Zero Inflated Negative Binomial

ZIP Zero Inflated Poisson

ACKNOWLEDGEMENTS

This study is a result of collaborative efforts involving many people who assisted me with technical, material, morale and emotional support. I would like to extend my sincere gratitude and profound thanks to the lecturers and staff of the Faculty of Science for their technical support and advice. Similarly, I wish to thank Chief Abraham Kanime from Windhoek City Police and Mrs Alexandrina Petrus, Constable Statistics Department of City Police and Emergency for providing me with data and assisting me to categorise crime into violent and property.

I also wish to acknowledge the help provided by all my friends. I thank them for their support and assistance rendered in proofreading my thesis and in their choice of suitable computer software for data analysis and interpretation of the results. My gratitude also goes to my mom and dad for their moral support.

I would like to express my deep gratitude to my academic supervisor Dr Isack Neema and co-supervisor Dr Lillian Pazvakawambwa for their tireless mentoring, modification and guidance in data cleaning as I progressed in my thesis writing that culminated in the production and presentation of this research work. I would like to thank Dr Opeoluwa Oyedele for her invaluable assistance through this study as well. Their guidance will remain a source of inspiration for my lifetime.

Finally, I wish to sincerely thank the leadership and management of the Namibia University of Science and Technology (NUST) for according me the opportunity to pursue this degree course. My special thanks and profound gratitude goes to the Head of Department of Mathematics and Statistics, Dr Onesmus Shuungula whose professional support enabled me to pursue my studies and spend some hours away from my office. You have been a wonderful team and I am proud to have been associated with you.

DEDICATIONS

This thesis is dedicated to my mom and daddy for their encouragement and support during this busy time.

God bless them all.

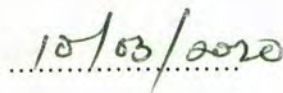
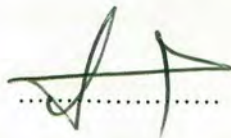
DECLARATIONS

I, Jonas Amunyela, hereby declare that this study is my own work and is a true reflection of my research, and that this work, or any part thereof has not been submitted for a degree at any other institution.

No part of this thesis/dissertation may be reproduced, stored in any retrieval system, or transmitted in any form, or by means (e.g. electronic, mechanical, photocopying, recording or otherwise) without the prior permission of the author, or The University of Namibia in that behalf.

I, Jonas Amunyela, grant The University of Namibia the right to reproduce this thesis in whole or in part, in any manner or format, which The University of Namibia may deem fit.

Jonas Amunyela,



Name of Student

Signature

Date

CHAPTER ONE

1. Introduction

1.1 Background of the study

Criminality originates from the Latin word “Crimen” which is defined as a social phenomenon that cannot be ruled out from the individual’s life (Khan, Ahmed, Nawaz, & Zaman, 2015). These scholars further argued that crime is a wrongdoing since it is an attempt to acquire something for nothing or the commission of an illegal act by law. An individual is less likely to be involved in criminal activities when there are substantial rewards and when he or she enjoys respect from the society to which they belong (Khan et al., 2015). In addition to that if a young person is gainfully engaged, either in education or employment, he or she is less likely to turn to crime (Dore, 2013).

According to the United Nations Office on Drugs and Crime (UNODC, 2010), it is particularly evident when one considers the recent trends and patterns of crime, especially, intentional homicide. This type of crime has a decreasing rate in Europe, Asia and Oceania (ranging between 2 to 4 per 100 000 population on average) and considerably higher rates in America (around 16 per 100,000 population, with much higher rates in the central and southern parts of America). In most European and African countries population density was found to be appropriate in crime prediction (Hanley, Lewis, & Ribeiro, 2016).

Tonry (2014) argued that the rate for Violent and Property crimes rose recently in all wealthy Western countries. Few years later, Property crimes in Western countries were declining as a result of improved security technologies in motor vehicles, residences, and retail stores (Tonry, 2014). This argument is also supported by the collaborative online database (NUMBEO, 2018) which enables users to share and compare information about crime index between countries and cities. The NUMBEO crime index indicates that in the Caribbean alone, the average crime index is currently standing at 71 % while the safety index is only at 29%.

Shaw and Reitano (2013) found that organised crime is closely linked to the development and changing nature of the African state itself and has been facilitated by the increasing connections between global economy and Africa. Crime has negatively affected several African countries and this led to the identification of three main phases of crime which occurred around tangible shifts in the prevailing social, political and economic environment (Shaw & Reitano, 2013).

Generally, crime erodes Africa's social and human capital by degrading the quality of life and pushing skilled workers overseas (Boyo, 2013). This has been termed as 'brain drain'. Khan et al., (2015) argued that the main interest of an economist is in the economic factors that determines the crimes in a society. Vogel and Van Ham (2018) contend that, the economy plays an important role as far as crime and social problems are concerned. Strong social structure, where all age categories have free access to education to improve their skills and knowledge is a basic requirement for economic growth (McGrory, 2018). Therefore, improving skills and knowledge for all age groups enhance employment opportunity.

Drug trading, kidnapping, embezzlement, the large scale theft of minerals, or other plainly criminal activities exist in many parts of Africa (Ellis & Shaw, 2015). In the Democratic Republic of Congo (DRC), such criminal activities seem to be most important if they concern high value natural resources, while in Mali, patterns of state protection for the illegal trade in cocaine were comparable to the experience of Sierra Leone and Guinea-Bissau (Ellis & Shaw, 2015). Palmary, Rauch, and Simpson (2014) contend that Johannesburg is believed to be the "crime capital" of South Africa. Beside that a total of 267 arrests related to the illegal rhino horn trade were made in South Africa for 2012. This was linked to the arrest figures of 165 for 2010 and 232 for 2011 (Ayling, 2013). It was found that the illegal wildlife networks operating in South Africa and Namibia had no one distinct profile but both poachers and traffickers of rhino horn tended to be informal groups or predominantly individuals (Ayling, 2013).

Namibia is situated in Sub-Saharan Africa, a region that has one of the highest crime rates in the world (Neema & Böhning, 2012). (Namibia Statistics Agency (NSA), 2011) estimated that the Namibian population was 2113077, and the (UNODC, 2010) calculated the rate of murders in 2012 at 17.2 per 100 000. This is relatively high when compared to other African countries. In comparison, Guinea-Bissau had an 8.4 murder rate per 100 000 people according to a UN survey data. Due to the constant increase in crime rate within Namibia, the former President Hifikepunye Pohamba suggested in 2014, that 6 March be declared a national prayer day.

Policymakers and concerned citizens in Namibia have introduced proper safety procedures that are efficient and effective to curb the crime rate. Currently, some regions have introduced an active neighbourhood watch and Operation Kalahari to curb residential break-ins. However, these operations sometimes fails to catch the thieves in the act.

A study by Sandema (2005) that examined the factors that influence armed robbery in Namibia, confirmed that there is a significant increase in gun crime incidents. It was established that young people between 18-30 years of age commit armed robbery (Sandema, 2005).

According to the Overseas Security Advisory Council (Council, 2010), Namibians have regularly fallen victim to street crime. Such incidents occur more frequently when it is dark. Criminals sometimes display knives and occasionally firearms. The most common incidents are non-violent crimes of opportunity which include, pickpocketing, purse snatching, vehicle theft, ATM card skimming, and vehicle break-ins. Pickpocketing and purse-snatching are most likely to take place in shopping centres as well as at high-traffic locations where foreign visitors to Namibia gather.

Motor vehicle theft remains a major concern. This type of crime usually involves smash-and-grab patterns and is sometimes associated with violence, especially when the occupants in the vehicle refuse to freely surrender their belongings to the perpetrators. Personal robberies and residential break-ins and thefts remain prevalent in Namibia. Notably, Windhoek City Police (WCPOLS, 2006), observe that ATM card skimming, purse-snatching, vehicle breaks-ins and vehicle theft are among the most frequently recorded Property crime types in Namibia. Card skimming is the act of using a skimmer to illegally collect data from the magnetic stripe of a credit, debit or ATM, without the owner of the card knowing. This data is later copied onto another blank card's magnetic stripe and used by the thief to make purchases or cash withdrawals in the name of the actual bank account holder.

The introduction of Operation Kalahari and the installation of CCTV cameras was aimed at crime reduction and prevention in Windhoek. The Operation Kalahari and its predecessor Operation Horncranz involved Namibian Police force, Windhoek City Police and Namibian Defence Force working together to achieve a common goal.

According to Neema and Böhning (2012), Karas and Khomas region experience a long-term increase in murder risk, which was considered as Violent crime in this study. As such, research indicates that crime continues to increase, especially in Windhoek. In addition, more than two fifths of all reported crimes in Namibia occur in Windhoek, where most of the reported crimes are burglaries, robberies and assaults (WCPOLS, 2006). The United State Statistics Department report (2015), noted that the other common type of crime in Windhoek is petty street crime. The Windhoek City Police has reported a slight increase in crime over the past years resulting in residents of the capital starting up neighbourhood watch groups as a security measure and defence mechanism.

- i) Assess the ecological characteristics of Violent and Property crimes in Windhoek with a view to create understanding of their root causes.
- ii) Evaluate Violent and Property crimes in Windhoek and determine locations with high crimes or crime hot spots for possible interventions.
- iii) Determine the season of the year in which Property and Violent crimes happen more often.
- iv) Model factors influencing Violent and Property crimes in Windhoek and assess their impact on the two crimes.

1.4 Significance of the Study

This study would contribute positively to the expansion of the knowledge base related to crime analysis in Windhoek. It would further offer guidance and information to policy makers, City police, government, and other interested organisations involved in crime control in Namibia. In addition, this study would assist in identifying areas in the City with high crime incidences for possible interventions.

1.5 Limitation of the study

Literature available on this study in Namibia was quite limited. For this reason, there is a shortage of adequate resources required to carry out this research and to consider the whole country for this study. In addition, due to hefty transport expenses, paperwork, limited access to the Chief Commanders of police across all cities in the 14 regions of Namibia and time limitation, the acquirement of primary data on crimes from each region within Namibia is a limitation for this study. Hence, the crime data considered for this study was secondary data received from Windhoek police, focusing only on Windhoek reported crimes. However, due to the sensitive nature of this study, the researcher was given limited information and access to the crime data received from Windhoek police station.

CHAPTER TWO

2 Literature Review

2.1 Introduction

Violent crime includes homicide (murder and manslaughter), aggravated assault, and robbery while Property crime includes burglary, larceny and motor vehicle theft (Chamberlain & Hipp, 2015). Homicide is a global problem. According to the United Nations Office for Drugs and Crime (UNODC, 2010) the highest homicide levels are found in America and the African region, with the lowest homicide levels generally in European countries.

Crime and insecurity are major challenges in African countries, threats to national development and individual quality of life (Wambua, 2015). In addition to that, only 11 African countries ranked in the top hundred countries worldwide in terms of safety and security with Benin being the top-ranked African country at No.50 (Legatum Institute, 2014). Subsequently, 38% Africans say they have felt unsafe walking in the neighbourhoods. South Africans are most likely to feel unsafe walking in their neighbourhoods and to fear crimes in their homes, while Niger, Ghana and Mauritius residents feel safest (Wambua, 2015).

Besides that, children under the age of five were second largest homicide victims after 15-19 years old in South Africa (Abrahams et al., 2016). The study by Eze (2016) found that incidence of passion killing was burgeoning in the locations of study and very rampant among youths; the victims were mainly women. Study by Asongu and Kodila-Tedika (2016) found that in most African countries the wave of crime could be addressed if fight against corruption is taken seriously by governments.

2.2 Consequence of Violent and Property crimes

Crime especially violence is prevalent worldwide, with the World Health Organisation (WHO) figures reporting that at least 1.6 million people die yearly as a result of violence (Garcia-Moreno, Jansen, Ellsberg, Heise, & Watts, 2006). Property and violent crimes create fear among communities. People from communities with high levels of crime tend to be more fearful of crime than people residing in areas with comparatively lower levels of crime (Breetzke & Pearson, 2014).

Crime has destructively affected African nations as it erodes the continent's social and human capital by degrading the people's quality of life and pushing expert workers overseas (Mazzitelli, 2007). In addition to that Mazzitelli (2007), contends that crime hampers access to employment and educational opportunities and discourages the accumulation of assets. Schepers (2017) asserted that social disadvantages affect the emergence of crime propensity and the criminogenic exposure of individuals.

2.3 Factors influencing Violent and Property crimes

Scholarly opinion within the geography of crime and spatial criminology studies concurs that crime is highly concentrated in certain areas due to specific factors within those areas (Breetzke & Pearson, 2014; de Melo, Matias, & Andresen, 2015). Moreover, spatial patterns of these concentrations differ across crime types. Among youth, factors that influence crime are unemployment and lack of education (Dore, 2013), own house mortgage (Jones & Pridemore, 2012), crime-specific detection rate and prison population (Han, Bandyopadhyay, & Bhattacharya, 2013).

According to Cohen and Felson (2016), Violent crime results from the convergence of three elements; namely, suitable target, motivated offender, and absence of capable guardians. This means that a suitable target is any person or commodity (such as money) that may invoke

criminal inclinations; a motivated offender is anyone with an inclination to commit crime, and a capable guardian is a person who can protect a target. Cohen and Felson (2016), further established that firearm homicide and firearm suicide are opposite activities but are equally violent crimes. The strongest associations with local crime rates are for income, number of alcohol outlets in an area and housing tenure structure (Livingston, Kearns, & Bannister, 2014). In the *Anatomy of Violence*, (Raine, 2013) dissects the criminal mind with a fascinating, readable, and far-reaching scientific expedition into the body of evidence that reveals that the brain is a key perpetrator in crime causation. His speculations are premised on the genetic perspective that the seeds of sin are sown at the beginning of a person's life, giving rise to an abnormal physiological functioning that cultivates crime. This essentially means that genetically, some children can be identified as criminals right from birth. On the other hand, McShane and Williams (1997), contest that the behaviours of individuals such as guardianship, and attractiveness affect the risk of being victimised.

2.4 Statistical modelling of Violent and Property crimes

Various methods have been applied in analysing crime data, namely, the fixed dynamic Generalized Method of Moment (Han et al., 2013); Kendall's trend test (Moffatt & Goh, 2013); error correction model (Janko & Popli, 2015); weighted least square regression (Jones & Pridemore, 2012); ordinary least square (OLS) regression model and a geographical weighted regression model (Song & Liu, 2013); ordinary least square and negative binomial regression (Linning, Andresen, & Brantingham, 2017); random effect model (Harrison, 2014; Papps & Winkelmann, 2000)

Schielzeth and Nakagawa (2013) advise that in most instances nested data structures are ubiquitous in the study of ecology and evolution, and such structures can be modelled appropriately by mixed-effects models. Mixed-effect models are widely known as hierarchical

or multilevel models in the social and medical sciences and they feature random effects that allow clustering of data in groups (Stegmueller, 2013). Schielzeth and Nakagawa (2013) confirmed that with mixed-effect models, two predictors can be fitted using the syntax for crossed effects even if the design is nested.

Schielzeth and Nakagawa (2013) found that “Grouping structures might arise from repeated measurements on the same individuals, but also from spatial or temporal structures, family structures, and social groups of individuals, just to mention a few”. Gelman and Hill (2014) contend that the individual observations are found at the lowest hierarchical level. Schielzeth and Nakagawa (2013) allege that in practical applications, variables are modelled as random effects if the main interest lies in estimating variances and as fixed effects if the primary interest is for estimating the mean.

As was pointed out by (Clark & Linzer, 2015), the drawback of fixed-effects models was that they require the estimation of a parameter for each unit. This may reduce the model’s power and increase the standard errors of the coefficient estimates. Clark and Linzer (2015) claim that the problem is degenerated when the within unit sample size is very small, as the unit effects alone may account for most of the variation in the dependent variable.

On the other side, Winter (2013), notes that there are several options when the residual plot indicates non-linearity. These are to perform a non-linear transformation of the response or shift interest to some non-linear models especially when dealing with categorical data. If there are stripes in residual plots, it may be relevant to consider non-linear models as alternatives. The study by Schielzeth and Nakagawa (2013), established that if the assumption made by random effects models are correct, then random effect would be the preferred choice because of its greater flexibility, generalizability, and its ability to model context, including variables that are only measured at the high level.

Furthermore, assumptions made by random effects models, including exogeneity of covariates and the normality of residuals are as reasonable as those made by the fixed effect models when the model is correctly specified. That means a well specified random effect model can be used to achieve everything that fixed effects models can achieve and more besides. Schielzeth and Nakagawa (2013) exemplify that with respect to time series data, random effects models performed well even when the normality assumption is violated.

The ecological model for understanding violent crime explores the relationship between the individual and contextual factors and considers violent crime as the product of multiple levels of influence on behavior (Krug, Mercy, Dahlberg, & Zwi, 2002). In addition, a high level of residential mobility (where people move away from their dwelling units so many times), heterogeneity (highly diverse population) and high population density are all examples of such characteristics and each has been associated with violence crime. Justus and Kassouf (2013) established that affluent areas attract more criminals such as theft and robberies due to the opportunities available to them.

In contrary, findings by Brooks et al. (2017a) were that count data can be analysed using a generalized linear mixed model when observations are correlated in ways that require random effects. They further claimed that attempting to fit the GLMM models via a glmmTMB package with a log Normal-poisson model and covariate-dependent zero-inflation led to convergence failure, and hence they substituted with a similar model (a Negative-binomial model). A comparison of the results showed that the deterministic methods (gam, glmmTMB and inla) were all fast; gam was fastest, because gam fitted a simpler model. While on the other hand the stochastic methods (MCMCglmm and brm) were about an order of magnitude slower (Brooks et al., 2017a).

The Generalized Linear Mixed Models is of significant. GLMM include a random effect model which may be considered in this study based on its ability to estimate covariates both within and between clusters, its ability to partition variance at multiple levels, it examines variation in effects across cluster and it is parsimonious. Similarly, the Poisson regression, Negative Binomial (NB) model and the Zero-Inflated Negative Binomial (ZINB) model were explored. These models were worth being explored since they are widely used on count data and produce reliable results (Harrison, 2014).

Bolker et al. (2009) acknowledge that count data with so many zero values cannot be made normal by transformation. Even when one succeed to transform the data, this transformed data might violate some statistical assumptions or limit the scope of inference (one cannot extrapolate estimates of fixed effects to new groups). GLMMs combine the properties of two statistical frameworks that are widely used in ecology and evolution study. These are the linear mixed models (which incorporate random effects) and generalized linear models (which handle non-normal data by using link functions and exponential family e.g normal, Poisson or binomial distributions) (Bolker et al., 2009). Bolker et al. (2009) further conclude that GLMMs are the best tool for analysing non-normal data that involve random effects.

2.4.1 Zero Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM)

Count data has been analysed using generalized linear mixed models especially when observations are correlated in a way that require random effects. However, count data are often zero-inflated, containing more zeros than would be expected from the typical error distributions used in GLMMs (Brooks et al., 2017b). For example, crime counts may be exactly zero for some months based on effective policing but may vary according to the negative binomial

distribution for an area with poor policing. In addition, the Zero-Inflation model which is part of ZINB-GLMM estimates the probability of an extra zero such that a positive contrast indicates a higher chance of absence (e.g. $\beta_i < 0$ means fewer absences in a variable i unaffected by the fixed variable) while $\beta_i > 0$ means higher abundances in a variable i unaffected by the fixed variable.

Bolker (2018); Brooks et al. (2017b), established a new R package, generalized linear mixed model Template Model Builder (*glmmTMB*), that increases the range of models that can easily be fitted to count data using maximum likelihood estimation. The interface is simply developed to be familiar to users of the *lme4* R package, a widely used tool for fitting GLMMs. All one must do, in principle, is to specify a distribution, link function, and structure of the random effects (Bolker et al., 2009).

The Template Model Builder (TMB) was known for maximising speed and flexibility through utilising automatic differentiation to estimate model slopes and the Laplace approximation for handling random effects. The strength of the R package *glmmTMB* lies on the number of benefits it poses. Among others, *glmmTMB* is more flexible than other packages available for estimating zero-inflated models via maximum likelihood, and faster than packages that use Markov chain Monte Carlo sampling for estimation (Brooks et al., 2017b).

Furthermore, it is also rated high in terms of flexibility for zero inflated modelling than INLA even though speed comparisons vary with model and data structure. Study results by Bolker et al. (2009) were that repeated measurements on the same individual, the same location, or observations taken at the same point in time are often correlated and this correlation can be accounted for using random effects in a GLMM.

There are five main packages in R available for modeling zero-inflated data: *pscl*, *INLA*, *MCMCglmm*, *glmmADMB* and *brms* (Brooks et al., 2017b). The package *pscl* can fit zero-inflated GLMs with predictor variables on the zero-inflation using maximum likelihood estimation even though it cannot model the correlation within individuals if they are sampled repeatedly (Zeileis, Kleiber, & Jackman, 2008). This phenomenon requires random effects. Omitting these effects and thereby ignoring correlation makes statistical tests anti-conservative (Bolker et al., 2009).

The *glmmADMB* package can fit zero-inflated GLMMs that contain random effects to account for correlation among observations but it cannot fit models with predictor variables in the zero inflation part of the model (Fournier et al., 2012). This package can only be appropriate for limited cases where all observational units have an equal probability of producing structural zero. INLA has the same limitation as *glmmADMB*.

The *MCMCglmm* and *brms* packages can fit zero-inflated GLMMs with predictors on zero-inflation, but they are relatively slow. Based on these limitations and challenges, the researcher preferred using the newly developed R package *glmmTMB* that can easily estimate zero-inflated GLMMs using maximum likelihood. The ability to fit these types of models quickly, using a single package will make it easier for the researcher to find the best model to explain patterns in the data. According to Berridge (2011), in GLMMs, the explanatory variables and the random effects (for a two level model, these are x_{ij} , z_j and μ_{0j}) affect the response (for a two-level model, this is y_{ij}) via the linear predictor (θ_{ij}), where:

$$\theta_{ij} = \gamma_{00} + \sum_{p=1}^P \gamma_{p0} x_{pij} + \sum_{q=1}^Q \gamma_{0q} z_{qj} + \mu_{0j}. \quad (2.1)$$

Let $i = 1, 2, \dots, m$ denote the index of the observational unit while $j = 1, 2, 3, \dots, m_i$ denotes the index for the within observation in this unit. In this context $\theta = [\beta^T, \sigma^T]^T (d \times 1)$ denotes the vector of model parameters, where $\beta_{(p \times 1)}$ represents the parameters for the fixed effect,

and $\sigma_{(s \times 1)}$ includes the parameters for the random effects while $d = p + s$. The observed outcome y_{ij} is assumed to be independently drawn from exponential family of distribution when conditioned on a vector $X_{ij}(p \times 1)$ and the random effects vector $\gamma_i(q \times 1) \sim N_q(0, \Delta_\sigma)$. In this case Δ_σ is a positive-definite symmetric covariance matrix. For simplicity, the study considered reparametrization $\gamma_i = D_\sigma u_i$ resulting from the Cholesky decomposition of $\Delta_\sigma = D_\sigma D_\sigma^T$ where u_i denotes the multivariate standard normal vectors. The GLMM is of the form:

$$g(u_{ij}) = \eta_{ij} = X_{ij}^T \beta + Z_{ij}^T D_\sigma u_i. \quad (2.2)$$

Where u_{ij} denotes the conditional expectation of the outcome, $Z_{ij}(q \times 1)$ a design vector for the random effects and η_{ij} the linear predictor. Furthermore, $g(\cdot)$ represents a link function which maps the linear predictor and the conditional expectation of the outcome (Flores-Agreda & Cantoni, 2019).

The GLMM is obtained by specifying some function of the response (y_{ij}) conditional on the linear predictor and other parameters, i.e.

$$f(y_{ij} | \theta_{ij}, \phi) = \exp \left\{ \frac{[y_{ij} \theta_{ij} - b(\theta_{ij})]}{\phi} + c(y_{ij}, \phi) \right\}, \quad (2.3)$$

where ϕ is the scale parameter, $f(\cdot)$ denotes the Probability Density Function (PDF), $b(\theta_{ij})$ is a function that gives the conditional mean (μ_{ij}) and variance of y_{ij} , namely:

$$E[y_{ij} | \theta_{ij}, \phi] = \mu_{ij} = b'(\theta_{ij}), \quad (2.4)$$

$$Var[y_{ij} | \theta_{ij}, \phi] = \phi b''(\theta_{ij}), \quad (2.5)$$

while $c()$ is a function that is automatically determined once the other functions have been chosen (or simply denotes a specific function), so that the entire distribution is normalized.

In Generalized Linear Mixed models, the mean and variance are related so that:

$$Var[y_{ij}|\theta_{ij}, \phi] = \phi b''(b'^{-1}(\theta_{ij})) = \phi V[\mu_{ij}]. \quad (2.6)$$

$V[\mu_{ij}]$ is referred to as the variance function, $b'^{-1}(\theta_{ij})$ is a link function which expresses θ_{ij} as a function of μ_{ij} , and $b'(\theta_{ij})$ is the inverse link function. The functions $b(\theta_{ij})$ and $c(y_{ij}, \phi)$ differ for different GLMMs. The distribution that works well in modelling the ZINB-GLMM is **nbinom2** by (Magnusson et al., 2017), which returns an overdispersion parameter.

The expressions of the marginal PDF are obtained after integrating the random effects from the joint distribution $[y_{ij}, u_i^T]^T$.

Since the study data was counts and zero inflated, the Poisson, Negative Binomial, Zero-Inflated Poisson, and Zero-Inflated Negative Binomial models were fitted as alternatives to the Zero-Inflated Negative Binomial Generalized Linear Mixed Model.

2.4.2 Poisson model

Poisson regression models have received much attention in econometrics literature as models for describing count data (Cameron & Trivedi, 2013). However, the Poisson regression model involves the restrictive assumption that the variance is equal to the mean. In this study, the count response variables Y_1 and Y_2 were number of Property and number of Violent crimes, respectively. The independent variables X_i , for $i = 1, 2, \dots, 5$ were Month, Year, Season, Location and Density.

The study fitted a Poisson regression model of the general form:

$$[E[Y_i/X_i] = \lambda_i = \exp(X_i' \beta_j) = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_5 X_5)] \quad (2.7)$$

Where $j = 1, 2, \dots, 5$, β_0 is the intercept and β_j are the coefficients.

2.4.3 Negative Binomial (NB) model

The Negative Binomial model has been used to model count data that are over dispersed. Overdispersion is common in models of count data in ecology and evolutionary biology, and can occur due to missing covariates, non-independent (aggregated) data, or an excess frequency of zeroes (Harrison, 2014). This model was more efficient than the Poisson model. However, NB models the probabilities and estimates the amount of extra variations. Following Zuu et al. (as cited in Harrison, 2014), the NB distribution can be parameterized as:

$$q(y|k, \mu) = \left(\frac{\Gamma(y+k)}{\Gamma(k) \times \Gamma(y+1)} \left(\frac{k}{\mu+k} \right)^k \times \left(1 - \frac{k}{\mu+k} \right)^y \right) \quad (2.8)$$

For $y = 1, 2, \dots, \infty$

Where the mean and variance of Y are respectively:

$$E(Y) = \mu \text{ and } var(Y) = \mu + \frac{\mu^2}{k}. \quad (2.9)$$

Thus, k controls the overdispersion level in the dataset. As k becomes large relative to μ , overdispersion is minimized and the model collapses to an ordinary Poisson distribution.

2.4.4 Zero-Inflated Negative Binomial (ZINB) model

A Zero-Inflated distribution for count data is a mixture of two distributions, the distribution that takes only the value zero; 'perfect state' and a distribution on the non-negative integers (i.e., including the value zero; 'imperfect state'). A sample is in the perfect state with probability p and the imperfect state with probability $q = 1 - p$. If the distribution for the

imperfect state is the Poisson, the mixture distribution is Zero-Inflated Poisson (ZIP). While on the other hand, if the distribution for the imperfect state is the Negative Binomial (NB), the mixture distribution is ZINB (Minami, 2007). The probability function for a ZINB regression model is expressed as:

$$f(y_i|B_i, G_i, \beta, \gamma, \theta) = p_i + (1 - p_i)q(0|\mu_i, \theta) \quad \text{for } y_i = 0 \quad (2.10)$$

$$f(y_i|B_i, G_i, \beta, \gamma, \theta) = (1 - p_i)q(y_i|\mu_i, \theta) \quad \text{for } y_i = 1, 2, \dots, \infty \quad (2.11)$$

Where $q(y_i|\mu_i, k)$ is given by (2.8). Covariates are related to the mean of the imperfect state and the probability of being in the perfect state as follows:

$$\log(\mu_i) = B_{i0} + B_{i1}\beta_1 + \dots + B_{ik}\beta_k = B_i\beta \quad (2.12)$$

$$\begin{aligned} \text{and} \quad \text{logit}(p_i) &= \log\left(\frac{p_i}{1-p_i}\right) \\ &= G_{i0} + G_{i1}\gamma_1 + \dots + G_{ik}\gamma_k \\ &= G_i\gamma \end{aligned} \quad (2.13)$$

Where B_i and G_i are row vectors containing covariate values for the i^{th} observation. Estimates for β, γ and θ are obtained by maximizing the log-likelihood function. The mean and variance for ZINB are given by:

$$E[Y] = (1 - p)\mu \equiv \mu^* \quad (2.14)$$

$$\begin{aligned} \text{and} \quad \text{Var}[Y] &= (1 - p)\mu + (1 - p)\left(p + \frac{1}{\theta}\right)\mu^2 \\ &= \mu^* + \frac{p + \frac{1}{\theta}}{1 - p}\mu^{*2} \end{aligned} \quad (2.15)$$

where θ is the dispersion parameter.

2.5 Model selection

The Likelihood Ratio (LR) tests can assess the significance of particular factors or, equivalently, choose the better of a pair of nested models, but Bolker et al. (2009) have criticised model selection via such pairwise comparison as an abuse of hypothesis. They further contend that, Akaike information criterion (AIC) and related information criteria (IC) use deviance as a measure of fit by adding a term to penalise more complex models (e.g. greater number of parameters). The AIC and residual plots performed well in model selection (Posada & Buckley, 2004).

2.5.1 Akaike Information Criteria (AIC)

The AIC is a popular method for comparing the adequacy of multiple, possibly non-nested models. Current practice is to accept a model with a small AIC value (Wagenmakers, 2004).

The equation of the AIC is described as;

$$AIC = -2LL + 2r \quad (2.16)$$

Where LL is a log likelihood value, r indicates number of parameter and n is a sample size (Posada & Buckley, 2004).

2.5.2 Residual plot for ZINB-GLMM

In assessing the best fit model, the study used the R package DHARMA by (Walker, 2018). DHARMA stands for “Diagnostics for Hierarchical Regression Models”. DHARMA includes support from glmmTMB and it is suitable for testing whether a GLMM is in harmony with the data (Dunn & Smyth, 1996). However, there are still a few limitations such as misspecifications in GLMMs which cannot be reliably diagnosed with standard residual plots. The expected

distribution of the data changes with the fitted values and that makes GLMM residual harder to interpret.

The current standard practice is to eye ball the residual plots for major misspecifications, potentially have a look at the random effect distribution and then run the test for overdispersion, but this approach still possesses a number of problems. Among these are, overdispersion which often comes from missing or misspecified predictors; not all overdispersion is the same and dispersion varies with the predictors (heteroscedasticity). DHARMA aims at solving these problems by creating readily interpretable residuals for Generalized Linear Mixed models that are standardised to values between 0 and 1. This is achieved through a simulation based approach, similar to the Bayesian p-value that transforms residuals to a standardized scale. The scaled (quantile) residuals are computed with the `simulateResiduals()` function in R. The default number of simulation ($n = 250$) was considered to be a reasonable deal between computation time and precision. What the function does is to:

Create n -new synthetic datasets by simulating from the fitted model, Compute the cumulative distribution of simulated values for each observed value and then return the scaled value that correspond to the observed values (Walker, 2018).

In short, the ZINB-GLMM sounds reasonable for this study because of its ability to handle multiple random effects components together with the zero-inflation and dispersion components. ZINB-GLMM performed well among other models for non-normal count data involving non-structural zeros due to its greater flexibility, generalizability, and its ability to model context, including variables that are only measured at high level. The model selection was done using AIC and Residual plots.

CHAPTER THREE

3. Research Methodology

This chapter outlines the research design, population, sample, research instrument and data analysis procedures. Subsequently, the ethical considerations followed in conducting this research were also articulated in this chapter.

3.1 Research Design

A quantitative design was adopted in this study. The research relied on secondary data on daily reported Property and Violent crimes obtained from the Windhoek City police (2011 to 2016).

3.2 Population

The population of this study is all Property and Violent crime cases reported in the City of Windhoek during the period 2011 to 2016.

3.3 Sample

No sampling was conducted since this study used data already available on reported Property and Violent crimes obtained from Windhoek City police.

3.4 Research instrument

The daily crime data were recorded in the pocketbooks by the City police officers who attended to crime scenes. After the pocketbooks were fully completed, they were submitted to the immediate supervisors or to the City Police Statistics Department for recording. The researcher obtained a letter from the University of Namibia requesting the crime data for a period of six years (2011 to 2016). The letter was presented to the Chief inspector of the City Police who then authorised the researcher to obtain the data. The soft copy of the data was obtained in a Microsoft excel document.

3.5 Data Analysis

In their annual report of 2016, the City police further categorised the reported crimes into Property crime and Violent crime. In this study, the response variables were the Number of Property crimes (Y_1) and Number of Violent crimes (Y_2).

The independent variables X_i , for $i = 1, 2, \dots, 5$, refer to the *Month*, *Year*, *Location*, *Season* and *Density*. The variable *Month* represents a specific month from January to December in which crimes were committed in a specific year within a location.

The variable *Location* represents fifty-nine Windhoek geographical areas namely, Academia, Babilon, Brakewater, Cimbebasia, Damara Location, Dolam, Donkerhoek, Dorado Park, Dorado Valley, Eehambo NdaNehale, Eros, Eros Airport, Freedom Square, Freedomland, Gemeente, Golgota, Goreangab, Green-Well Matongo, Grysblock, Hakahana, Havana, Herero Location, Hockland Park, Independence Arena, Khomasdal, Kilimajaro, Kleine Kuppe, Klein Windhoek, Lafrenz, Ludwigsdorf, Marua Mall, Malaka Draai, Maroela, Mix, Mukwanangombe, Mukwanekamba, Northern Industrial, Nubuamisi, Okahandja Park, Okuryangava, Olympia, Ombili, One-Nation, Ongulumbashe, Oshitenda, Otjomuise, Pionierspark, Properita, Rocky Crest, Shandumbala, Single Quarter, Southern Industrial, Soweto, Suiderhof, Vambo Location, Wanahenda, Windhoek Central, Windhoek North, and Windhoek West.

Prior to fitting the models for this study, a new variable *Season*, which represents all four seasons in a year in the Namibian context, was created. To be precise, Summer (January-March), Autumn (April-June), Winter (July-September), and Spring (October-December) (Kemper & ROUX, 2005).

The variable *Density* was obtained as the population per kilometer square of the area. The density was then scaled per 10 000 people since it was represented as large values which was

against the rule of thumb. A good rule of thumb is that input variables should be small values, probably in the range of 0-1 or standardized with a zero mean and a standard deviation of one. The dataset was cleaned and re-coded prior to the data analysis. Table 2A in the appendixA shows the codes assigned to all five independent variables.

Cross tabulation was carried out using the Statistical Package for Social Sciences (SPSS) version 25 (Green & Salkind, 2016). This was done to obtain the final counts of reported cases across the Property and Violent crimes. Table 3-1 below shows a count summary of seasonal reported cases under Property and Violent crimes, reported from 2011 to 2016.

Table 3-1: Summary of crime statistics across seasons and years

		No of Violent crimes		No of Property crimes	
		count	percent	Count	percent
Season	Summer	8063	31.3	17671	68.7
	Autumn	9430	31.1	20848	68.9
	Winter	9677	31.29	21248	68.7
	Spring	10633	32.8	21830	67.2
Year	2011	4498	25.8	12922	74.2
	2012	8862	37	15105	63
	2013	1288	66.9	637	33.1
	2014	7888	31.3	17343	68.7
	2015	8019	31	17970	69.1
	2016	7248	29.1	17620	70.9

There were 37,803 Violent crimes in total over six years: 4498 in 2011, 8862 in 2012, 1288 in 2013, 7888 in 2014, 8019 in 2015, and 7248 in 2016. Furthermore, there were 81,597 Property crimes in total over the study period: 12922 in 2011, 15105 in 2012, 637 in 2013, 17343 in 2014, 17970 in 2015, and 17620 in 2016. In all years with the exception of 2013, Property crime was higher than Violent crime. The percentages of Violent crimes were calculated by

dividing Violent crimes for each year by the total number of crimes for that year. The Property crime percentages were calculated in a similar manner.

There were 37,803 Violent crimes in total over the four seasons: 8063 in summer, 9430 in autumn, 9677 in Winter and 10633 in spring. Notably, of the 81,597 Property crime cases during the study period: 17,671 were counted in summer, 20,848 in autumn, 21,248 in winter and 21830 in spring.

3.6. Research Ethics

The City police chief authorised the researcher to use the data. The research commenced after the ethical clearance by the UNAM Research Ethics Committee (UREC), and research permission from the Centre for Postgraduate Studies were granted.

CHAPTER FOUR

4. DATA ANALYSIS AND PRESENTATION OF THE RESULTS

4.1 Introduction

This chapter presents the results and findings of the study. Chapter three conceptualised the research design and spelt out the methodologies that the study used to address the research problem. In this chapter, the researcher presents the data obtained from the Windhoek police. The aim was to bring out the views on and to discuss the issues pertaining to Property and Violent crimes in Windhoek. To accomplish the task of this research, the researcher relied on the police data of 2011 to 2016.

To be specific, this chapter present the outcome of the study in relation to the research objectives outlined in Chapter one. Findings show that crime statistics are only indications of those cases known to the police, as not all crimes were reported.

This research was conducted in Windhoek, the capital city of Namibia with a population of approximately 400 000 people. The Windhoek police Department divides the city into nineteen police zones, and these zones control fifty-nine locations (residential and non-residential). This study used crime data from all the fifty-nine locations since the data was the only one in the city with a long period of Property and Violent crime records, and with geographical data, which provides more stable estimates of the two crime cases. Based on police information, it can be assumed that these locations are representative of the city.

The study explored the generalized linear mixed model, Poisson, Negative Binomial, Zero Inflated Poisson (ZIP) and the Zero Inflated Negative Binomial (ZINB) modelling with R, an open source language and environment for statistical computing (Team, 2013). The lme4 package by (Kuznetsova et al.,2017) offers fast and reliable algorithms for parameter estimation for the mixed effect.

In the appendix A, there is Table 1A which present the dataset used in the research. This table has an initial header line specifying column names. The data is loaded into R simply as csv file and stored as an object *data1*. Four of the explanatory variables (Year, Location, Month and Season) were treated as factors in this study.

4.2 Data Presentation

Data exploration was carried out to provide information about the nature of the dataset. For variables that are continuous, the relevant summary statistics are shown and for categorical, a frequency tabulation is displayed.

The number of reported Property crimes ranges from 0 to 258 cases with an average of 41 cases per year. On the other hand, the number of reported Violent crimes ranged from 0 to 239 cases with an average of 19 cases per year. More Property crimes occurred in Windhoek as compared to Violent crimes. It can be clearly stated that the Number of Property crimes are twice as likely to occur as the Number of Violent crimes.

In the context of this study, Density is defined as the number of people per square kilometer of an area while Density2 is the number of people per ten thousand kilometers of an area. The variable Density was scaled into Density2 in order to avoid convergence issues in the models.

Finding shows that for the study locations, the minimum number of people per square kilometer is 41 while the maximum is 21 812. In addition, the average number of people per study locations was found to be 4611.

To obtain the first overview of the dependent variable (Number of Property & Number of Violent crimes), a histogram, boxplot and the normal Q-Q plot of the observed count frequencies were presented. Multiple plots of the Number of Property and Number of Violent crimes were displayed to check the crime pattern on a yearly, seasonal and monthly basis via

boxplots. The study further assessed the patterns of these crimes across all the fifty-nine (59) locations.

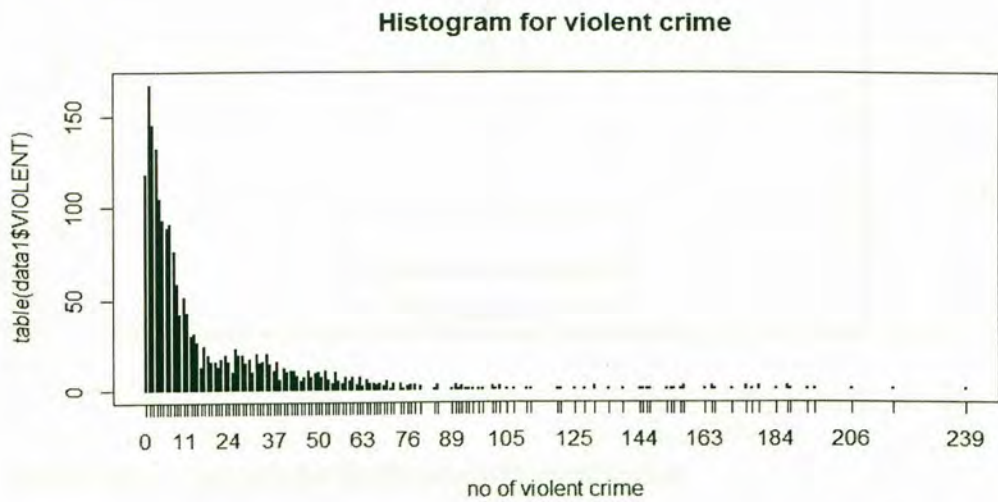


Figure 4.1a: Histogram for the Number of Violent crimes

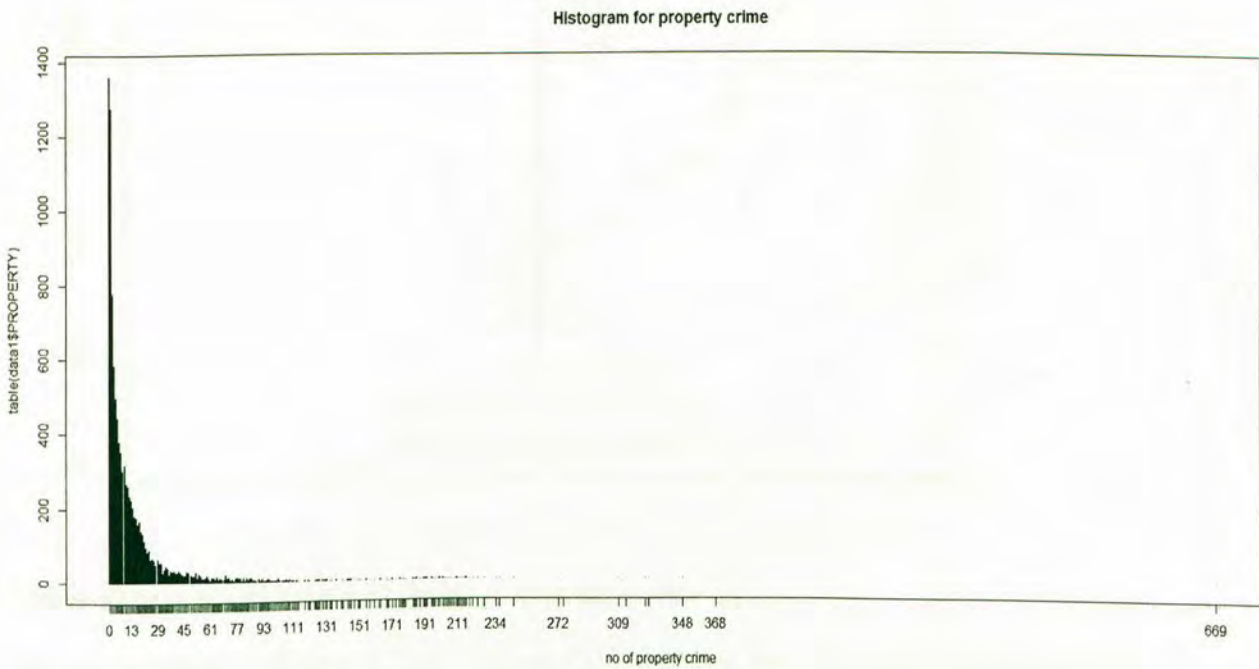


Figure 4.1b: Histogram for the Number of Property crimes

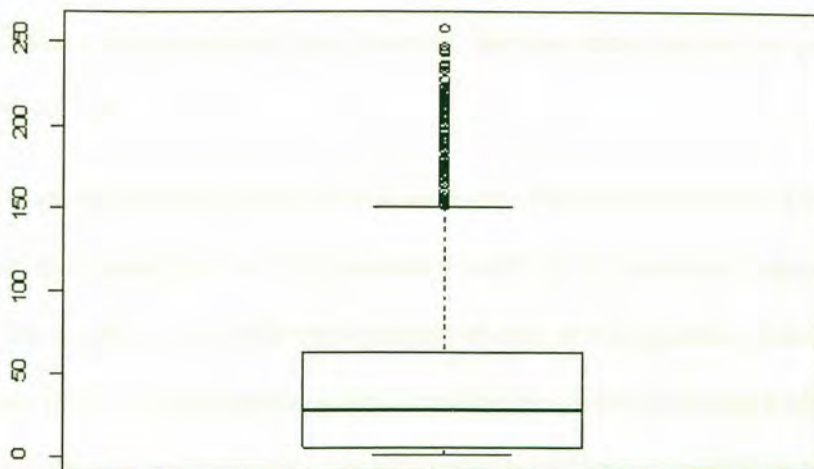


Figure 4.2a: A box plot for the Number of Violent crimes

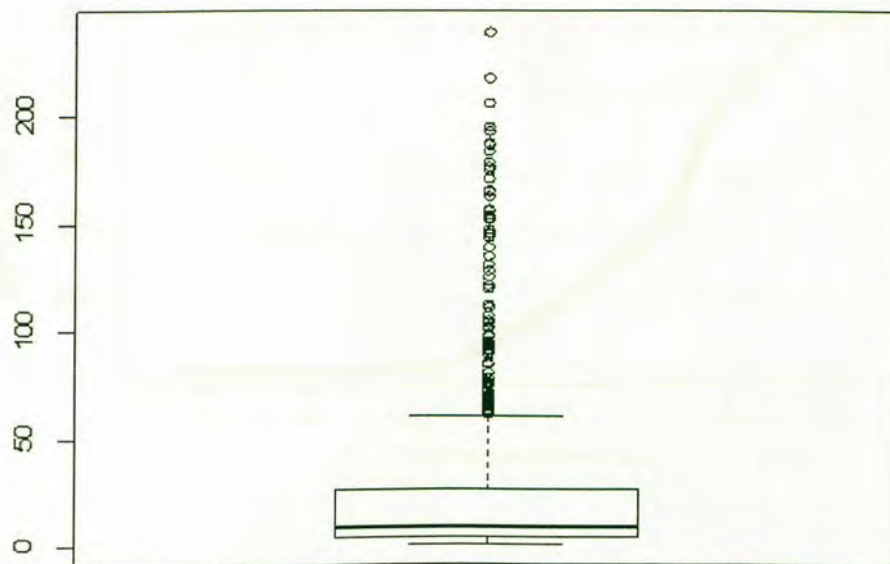
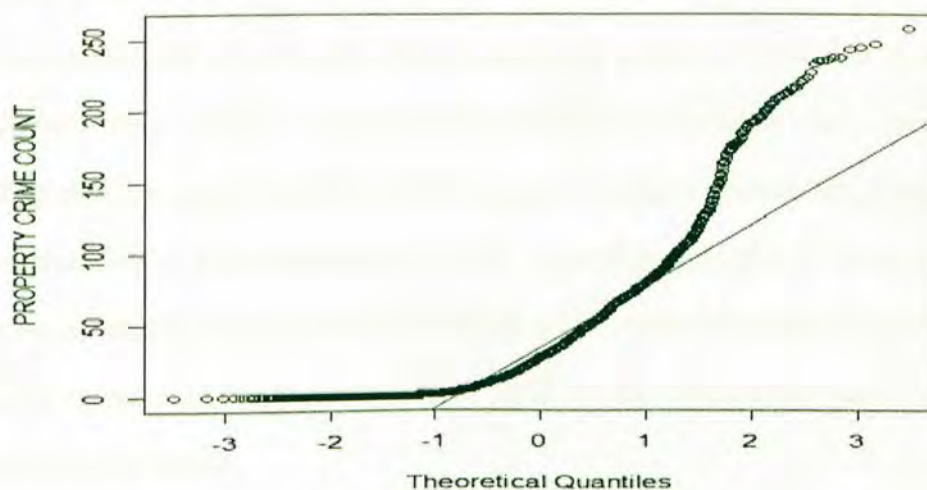


Figure 4.2b: A box plot for the Number of Property crimes

The two histograms (Figure 4.1a and Figure 4.1b) illustrate that the marginal distribution exhibits both substantial variation and a rather large number of zeros. It is clearly evident that the Number of Property and the Number of Violent crimes appears to be positively skewed, as

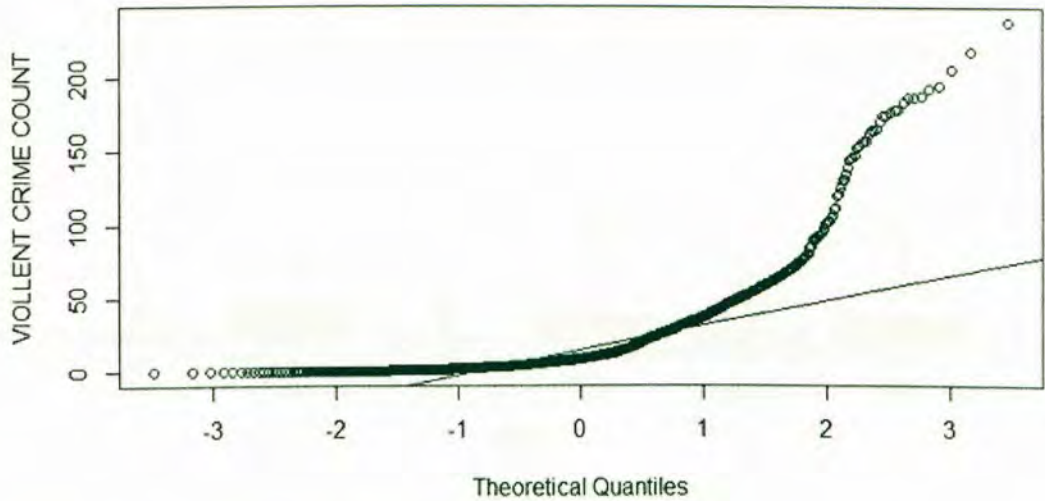
indicated by the relative position of the median within their box plots (Figure 4.2a and Figure 4.2b) that contains half the data. However, there are some few outliers as shown in the Figure 4.2a and 4.2b.

There are two distinct process driving zeros, one is non-structural zeros (sampling zeros) which occur by chance and can be assumed a result of dichotomous process. The other one is structural zeros (true zeros) which are part of the counting process. Based on this concern, the choice should be based on the model providing the closest fit between observed and predicted values. The choice of the zero-inflated model in this thesis is guided by the researcher's belief about the source of the zeros.



Shapiro Wilk normality test $W = 0.80677, p - value < 0.000$

Figure 4.3a: Number of Property crime Q-Qplot.



Shapiro Wilk normality test $W = 0.64277, p - value < 0.000$

Figure 4.3b : Number of Violent crime Q-Q plot.

The Q-Q plots show that both the Number of Property and Number of Violent crimes are positively skewed (not normally distributed) as the points fall above the line as x-values increases. This violates a very important assumption for the linear mixed effect model and rather considers generalized linear mixed models. Literature outlined that, linear models are not appropriate in some situations where the response is restricted to binary and count. In addition, linear models fail when the variance of the response depends on the mean. Finally, Shapiro's test indicates that it is safe to reject the null hypothesis, which states that the distribution is normal.

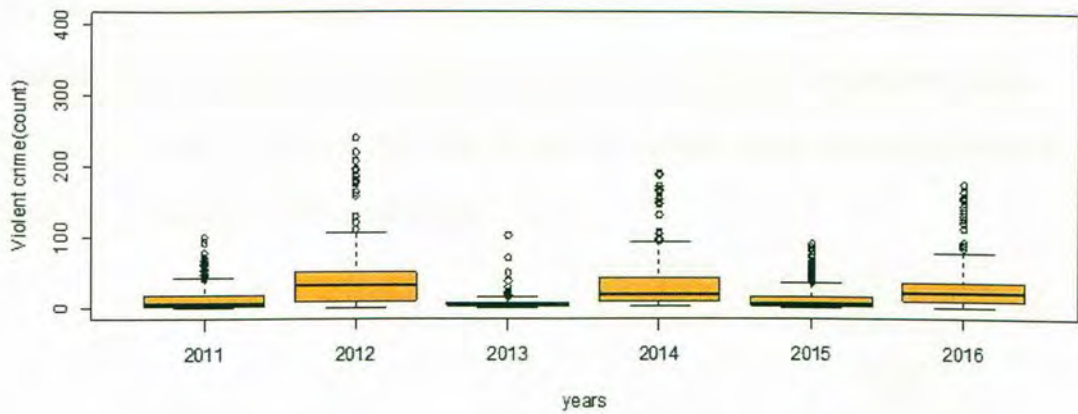


Figure 4.4a: Number of Violent crime distribution for each year

The boxplot presented in Figure 4.4a shows the variation in the average number of Violent crimes across the study periods, with the years 2012, 2014 and 2016 showing a very similar average of the number of cases. The years 2011 and 2015 also show a similar average of the number of cases. This indicates that the variation across the years needs to be taken into account when fitting the model.

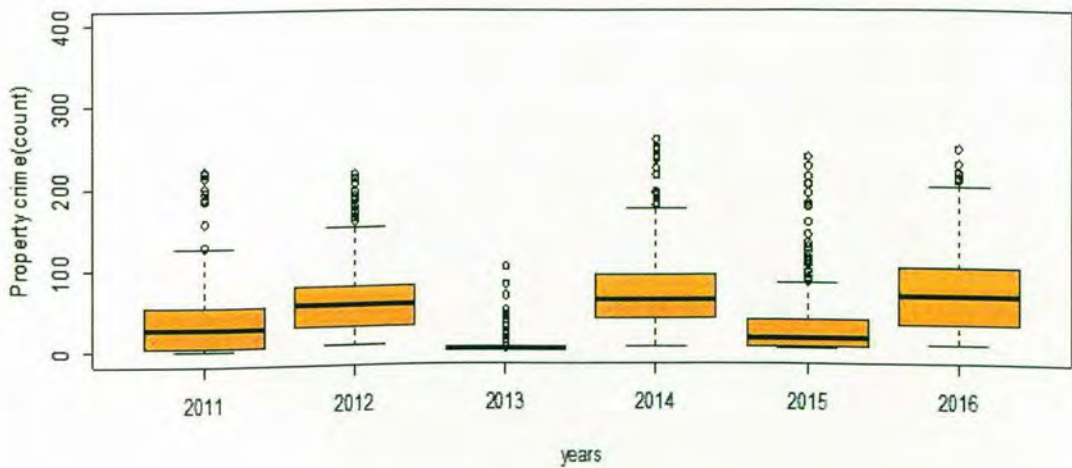


Figure 4.4b: Number of Property crime distribution for each year

The boxplot presented in Figure 4.4b shows the variation in the median Number of Property crimes across the study periods, with the years 2012, 2014 and 2016 showing very similar median of the number of cases. This indicates that the variation across the years needs to be taken into account when fitting the model.



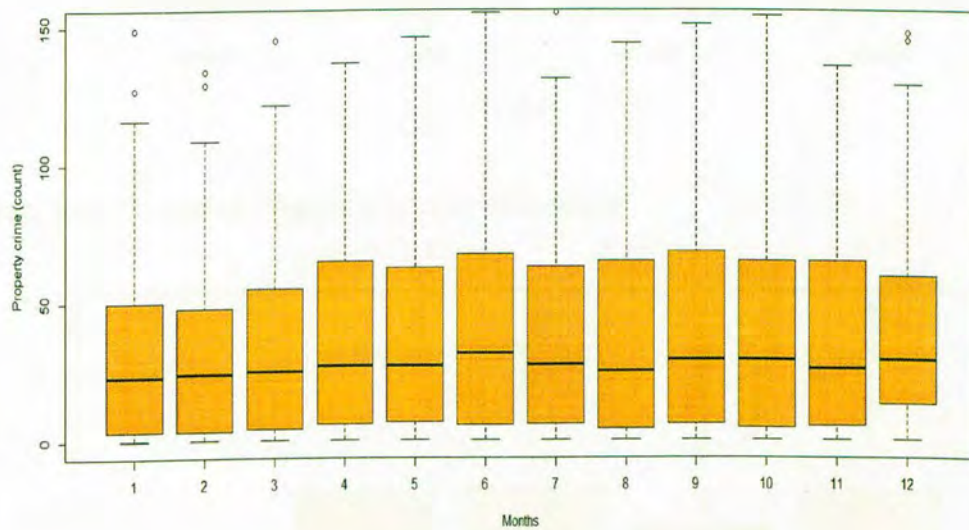


Figure 4.5: The Number of Property crime distribution for each month (January to December).

The boxplot shows that the median for the Property crime is slightly different. It also shows that each month presents a different amount of variation in Property crime so that there is an overlap of values between some months. The variation seems to be high as from April to December. There are still noticeable differences and hence the study better accounts for them in the model.

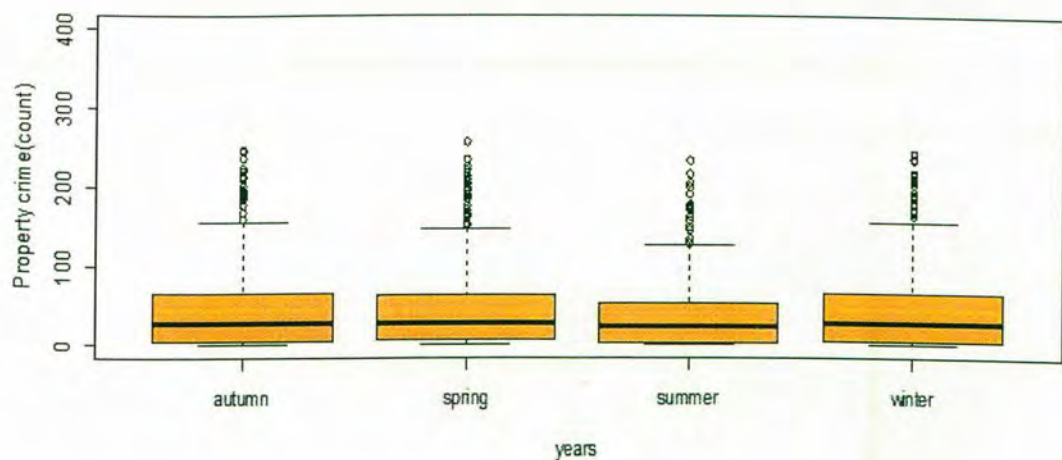


Figure 4.6a: Number of Property crimes for each season

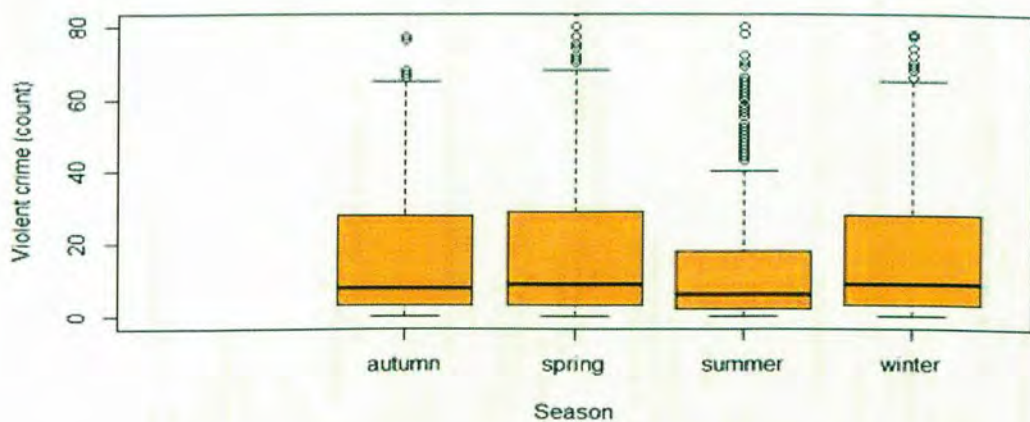


Figure 4.6b: Number of Violent crimes for each season

The above Figures 4.4, 4.5 and 4.6, demonstrate the possible presence of outliers in the study of the two types of crimes. Hence, suggesting that the study should not heavily rely on the average of Property and Violent crimes as they may distort the results. It was noted that in all the box plots (Figures 4.4, 4.5, 4.6a and 4.7b), the average Number of Property and the average Number of Violent crimes were above the median since the data was positively

skewed. Furthermore, the median for the Number of Property crimes was high when compared to the median of the Number for Violent crimes for all seasons, and so was the average. Therefore, the variation in Seasons was accounted for in the model.

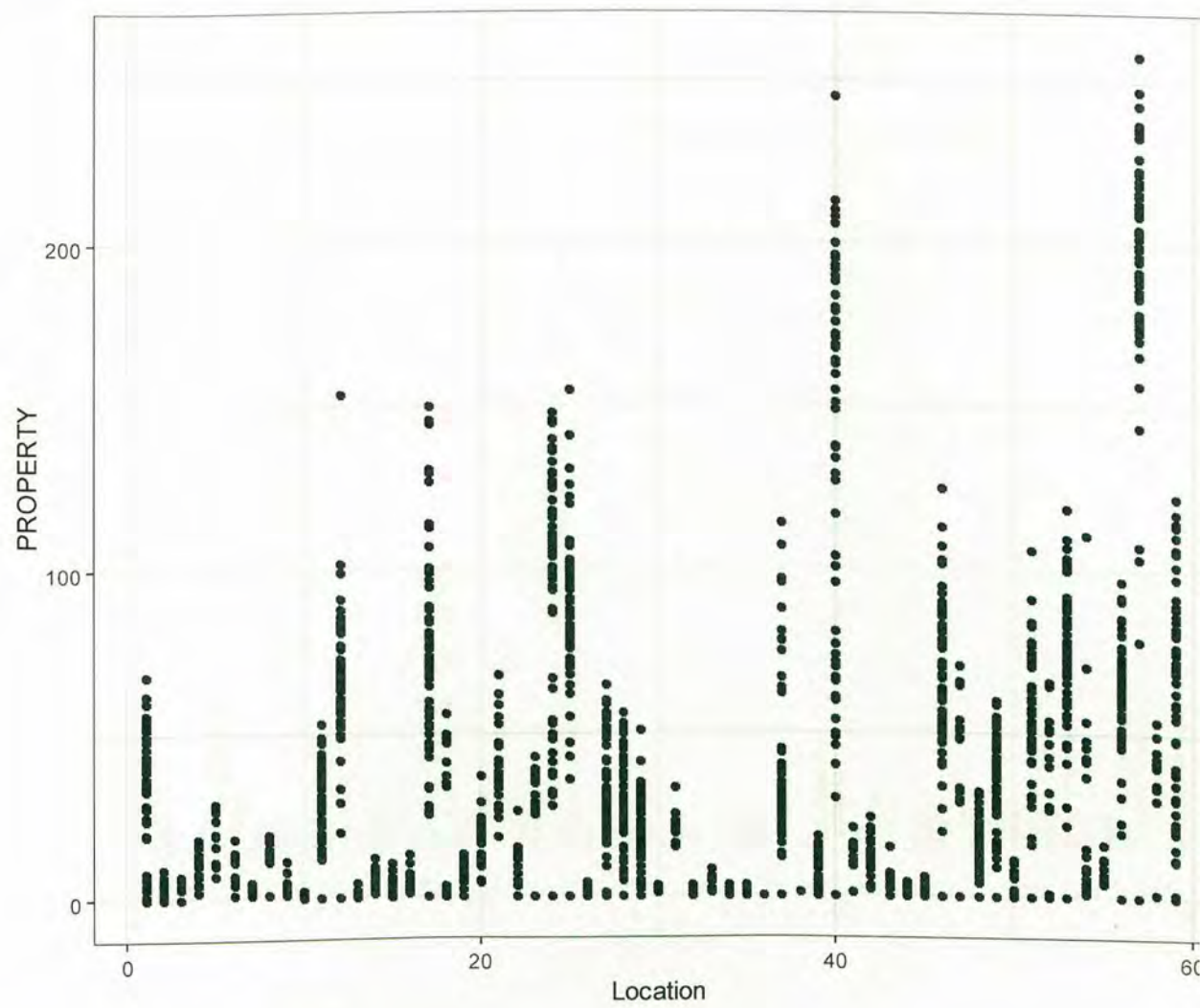


Figure 4.7a: Plotting Number of Property crimes against locations

From the above plot (Figure 4.7a), one can visualise that there is considerable variation among and within locations. The Number of Property crimes was high in some locations with a few outliers observed. To be specific Property crime was very high in Okuryangava and Windhoek central (location 40 and 57 respectively). In addition, Okuryangava has the second highest population density and that might constitute to crime incidents. In contrary, Windhoek central

was among locations with low population density but was more prone to crime as people gather around town every day.

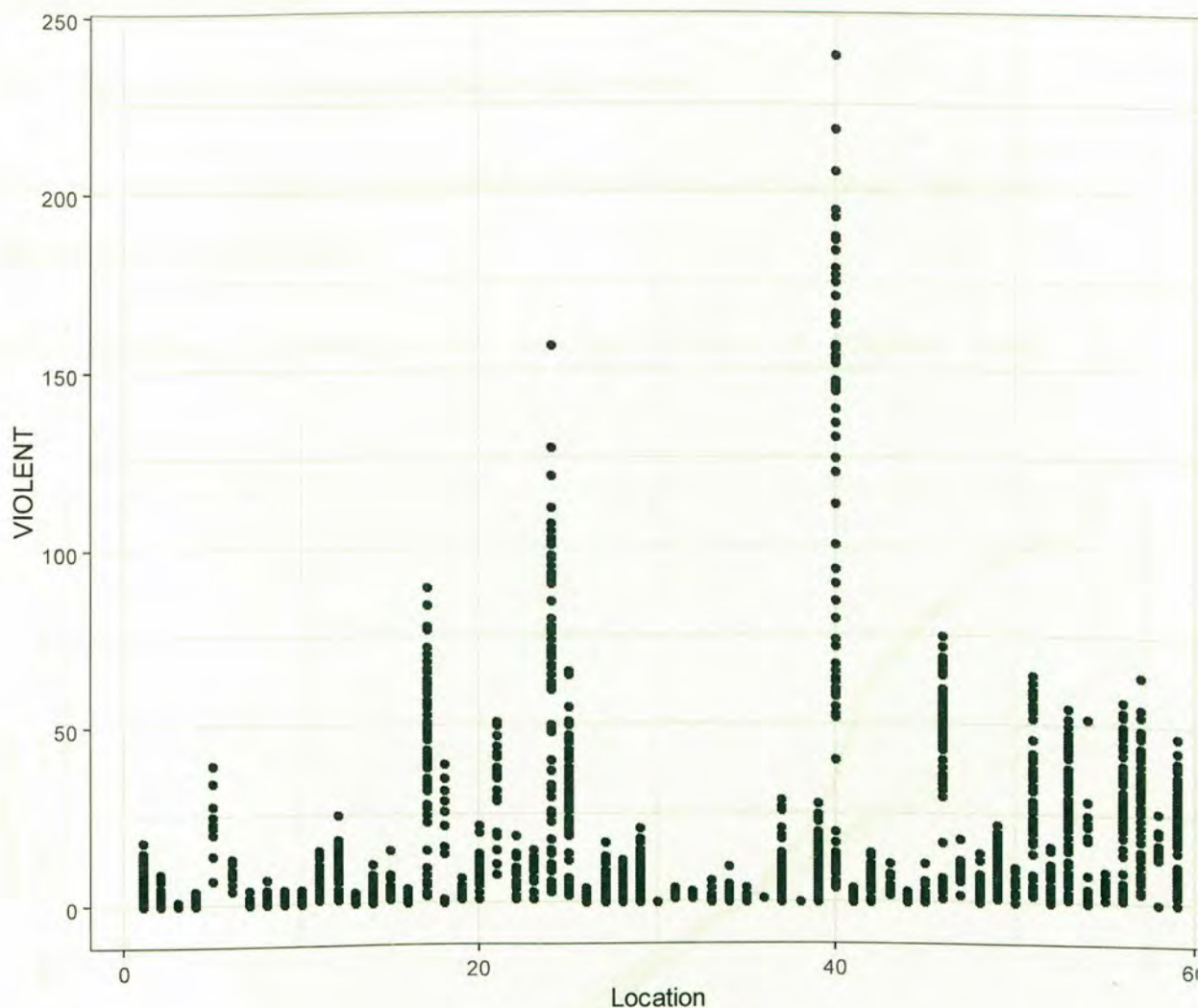


Figure 4.7b: Plotting Number of Violent crimes against locations

From the above plot (Figure 4.7b), one can visualise that there is considerable variation among and within locations. The Number of Violent crimes was high in some locations with a few outliers observed. It was found that Violent crime happened more often in Goreangab, Independence Arena, Okuryangava and Windhoek Central (location 17, 24, 40 and 57 respectively). Beside Windhoek Central, these locations were among locations with high

population densities. Even though Hakahana has the highest population density in Windhoek (21 594 people per km^2), the number of Property and Violent crimes were very low as compared to other locations.

4.3 Specifying the suitable model that best fits the data

The study considered finding a best model for this crime data. There are many ways to test this, but this study employed one.

4.3.1 Specifying a suitable model for the Number of Property crime.

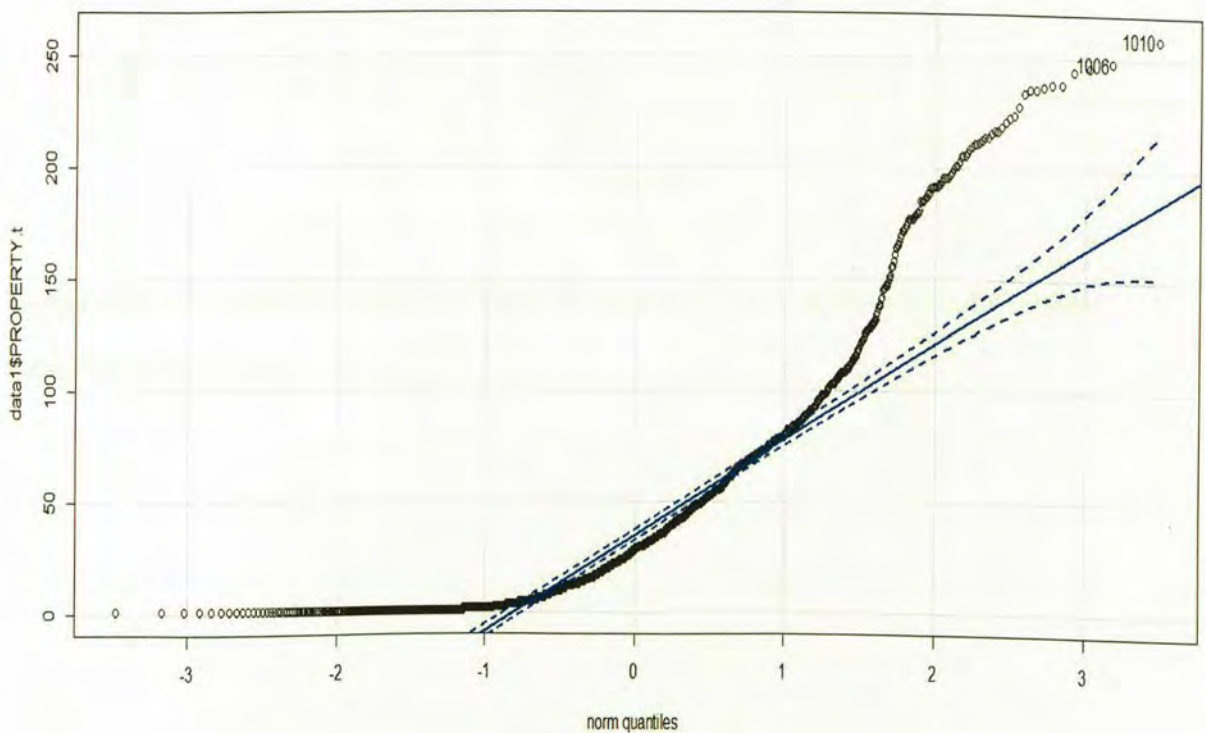


Figure 4.8a: Assessing if the Number of Property crime data follows a normal distribution using confidence interval

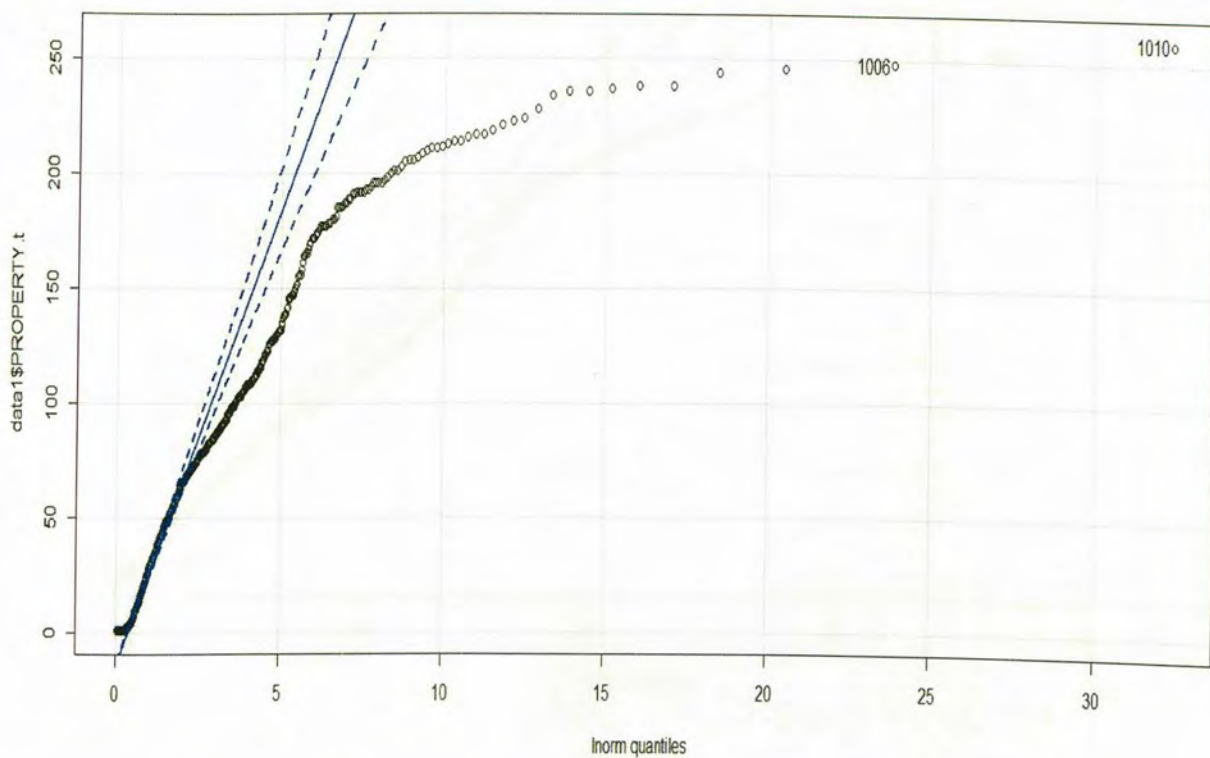


Figure 4.9a: Assessing if the Number of Property crime data follows a log-normal distribution using confidence interval

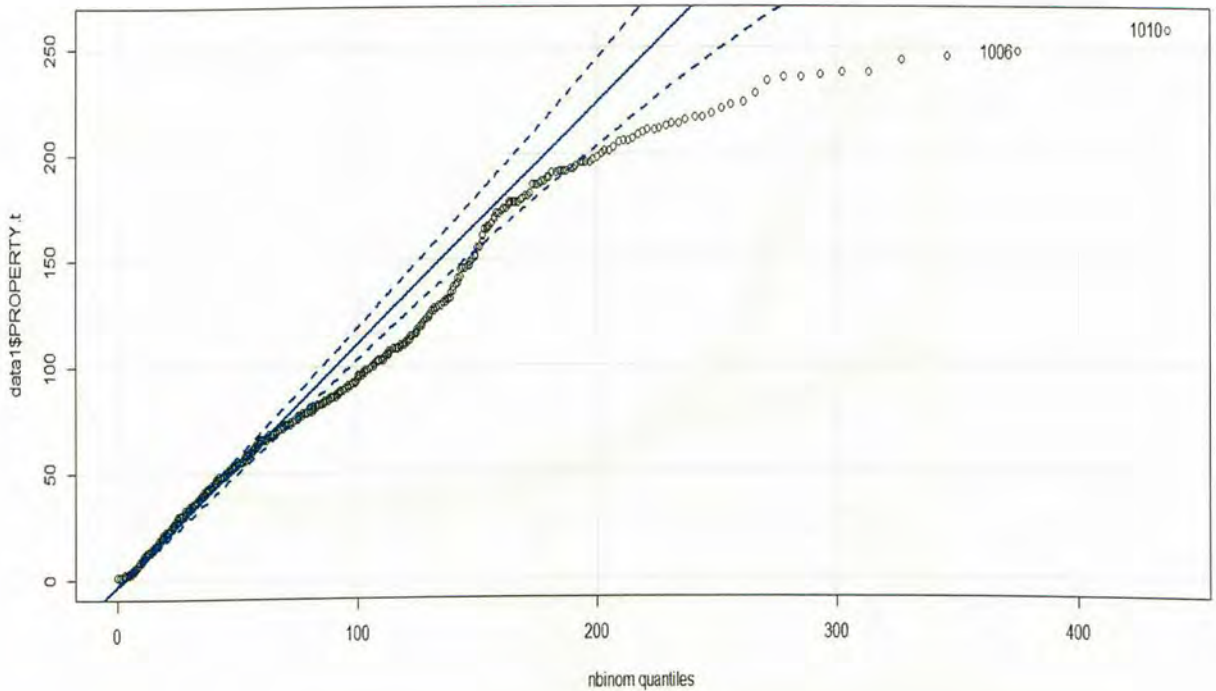


Figure 4.10a: Assessing if the Number of Property crime data follows a negative binomial distribution using confidence interval

Considering the plots generated using Q-Q plot, the y-axis represents the observations and the x-axis represents the quantiles modelled by the distribution. The solid blue line represents a perfect distribution fit and the dashed blue lines are the confidence intervals of the perfect distribution fit. The aim is to see if the data follows a normal distribution or other distributions. In this case, it is the negative binomial distribution, in which only a few observations fall outside the dashed lines. Now, armed with the knowledge that a negative binomial probability distribution fits best on Property crime data.

4.3.2 Specifying a suitable model for the Number of Violent crimes

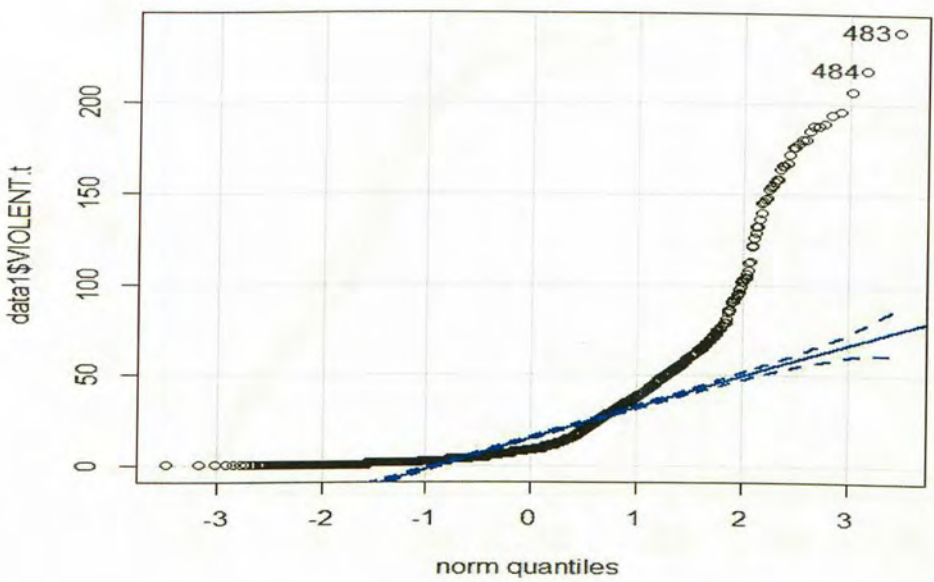


Figure 4.8b: Assessing if the Number of Violent crime data follows a normal distribution using confidence interval.

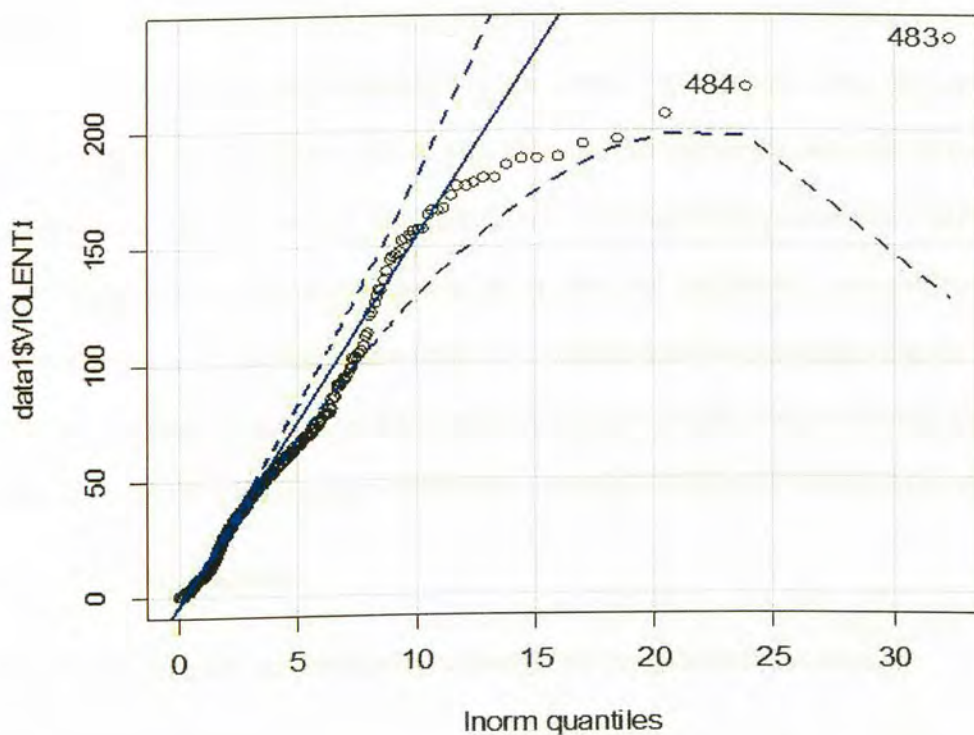


Figure 4.9b: Assessing if the Number of Violent crime data follows a log-normal distribution using confidence interval.

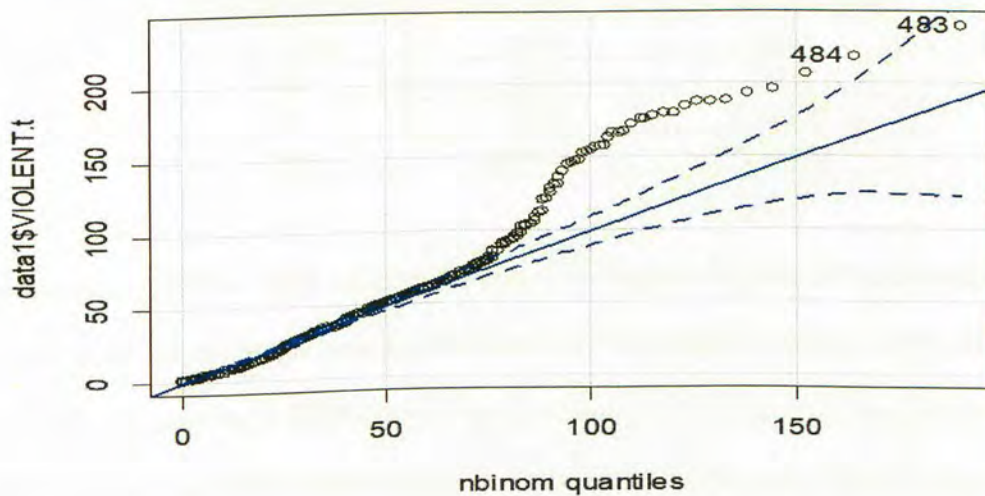


Figure 4.10b: Assessing if the Number of Violent crime data follows a negative binomial distribution using confidence interval.

Considering the plots generated using Q-Q plot, the y-axis represents the observations and the x-axis represents the quantiles modelled by the distribution. The solid blue line represents a perfect distribution fit and the dashed blue lines are the confidence intervals of the perfect distribution fit. The aim is to see if the data follows a normal distribution or other distributions. In this case, it is the log normal and negative binomial distribution in which only a few observations fall outside the dashed lines. Now, armed with the knowledge that the negative binomial distribution fits best on the Number of Property and the Number Violent crime data. It is also evident that log normal distribution could fit the Number of Violent crime data.

4.4 Variable selection.

Variable selection was not carried out because of the complexity of the model.

4.5 Model selection

Table 4-1: Comparing AIC for five Property and Violent crime models.

MODELS	PROPERTY CRIME	VIOLENT CRIME
ZINB-GLMM	14292.6	12343
POISSON	95365	55566
NB	18133	15215
ZIP	74613	52412.25
ZINB	17972.54	15210.02

In this study, ZINB-GLMM was reasonable in modelling the Number of Violent and the Number of Property crimes because of its small AIC values as compared to ZINB, ZIP, NB and Poisson (Table 4-2). ZINB-GLMM was the best model for this study based on the benefit that it accounts for within variation through random effects and captures the non-structural zero counts in the dataset.

Figures 4.11a and Figure 4.11b display the DHARMA scaled residual plots of ZINB-GLMM for Property and Violent crime, respectively.

DHARMA scaled residual plots

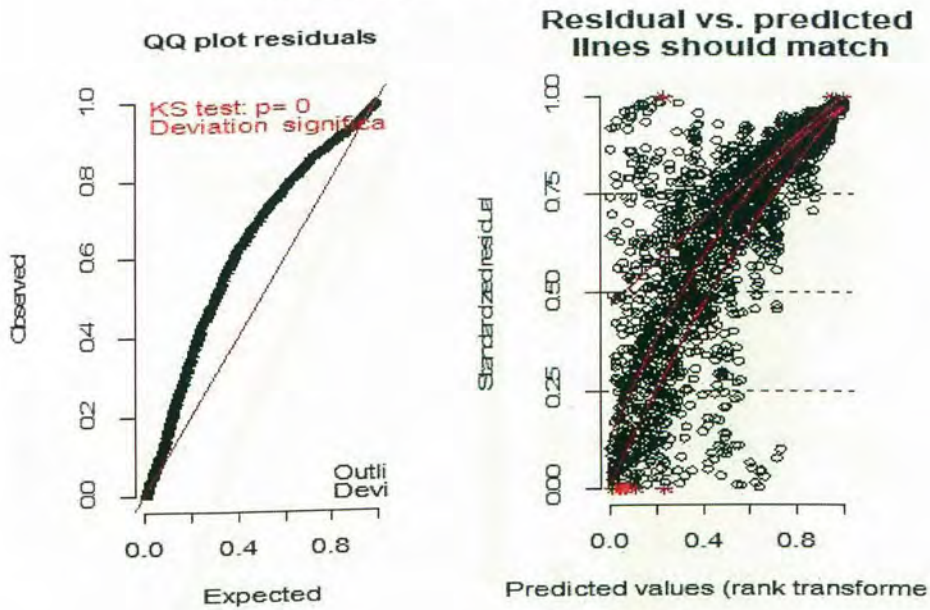


Figure 4.11a: Residual plot for Number of Property crimes model (ZINB-GLMM).

Figure 4.11a illustrates a better residual fits. There seems to be many residuals around 0.5, which means that the study is not getting as many residuals as expected in the tail of the distribution than expected from the fitted model.

DHARMA scaled residual plots

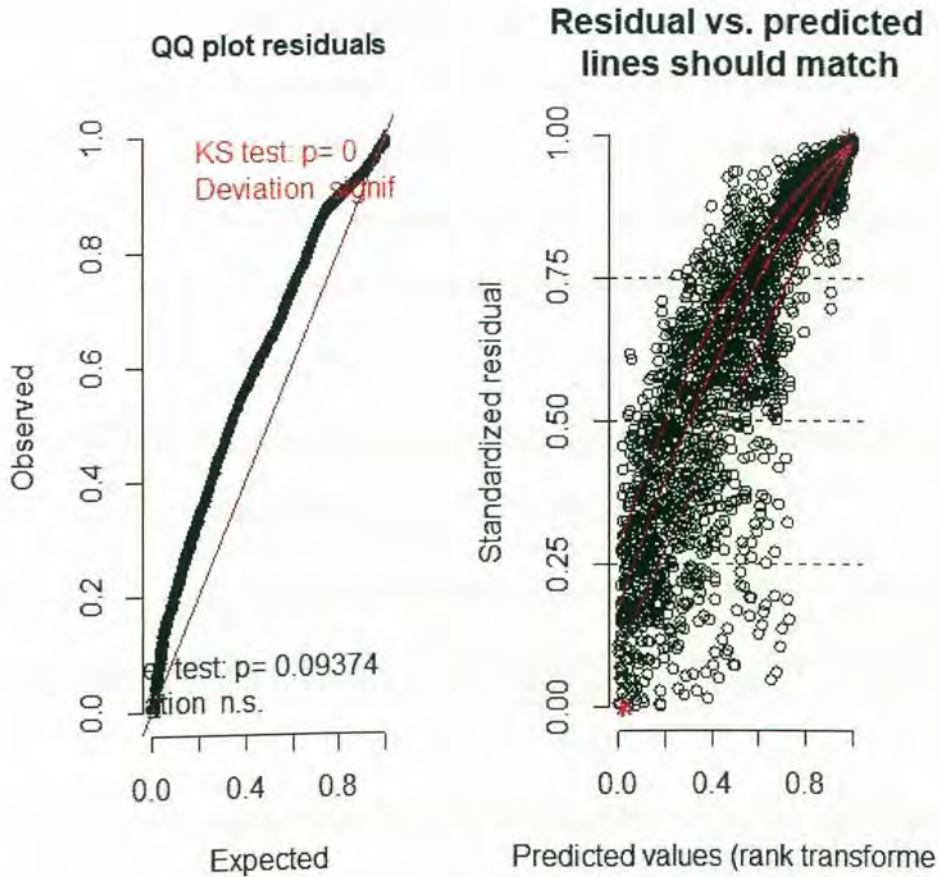


Figure 4.11b: Residual plot for Number of Violent crime model (ZINB-GLMM).

Based on Figure 4.11b, there are many residuals around 0.6, which means that the researcher is not getting as many residuals as expected in the tail of the distribution as well. The residual fit is not too convincing for this model but based on AIC the ZINB-GLMM, is the best.

4.5.1 Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) for the Number Property crime.

For the ZINB-GLMM, the response variable of interest was the Number of Property crimes while the independent variables were *Month*, *Season*, *Location*, *Year* and *Density*. The first

four variables (Month, Season, Location and Year) were the random effects. These effects were chosen to be “random” because the crime committed in one month is haphazard of the crime committed in the next month. This is also applicable to Location, Season and Year. However, Density was chosen to be the fixed effect since the number of people per square kilometer could be measured during this study period. Moreover, the random effects were nested in this study, because each police officer recorded a certain number of cases, and no two officers recorded the same case.

The full Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) is modelled on the familiar output format of lme4. The model fitted is:

$$\begin{aligned} \text{Mixedglmm1} < -\text{glmmTMB}(\text{PROPERTY} \sim \text{Density2} + (1|\text{Month}) + (1|\text{Year}) + \\ (1|\text{Location}) + (1|\text{Season}), \text{zi} = \sim \text{Year} + \text{disp} = \sim \text{Season}, \text{data1}, \text{family} = \text{nbinom2}) \end{aligned} \quad (4a)$$

Following the arguments in the function (Mixedglmm1), this model allows the conditional mean to depend on Density and vary randomly by Location (l), Month (m), Year (y) and Season (s). It further allows the number of structural (extra) zeros to depend on years. Additionally, it allows the dispersion parameter to depend on the season of the year. This model can be represented by the following specific equations (Brooks et al., 2017b).

$$\mu = E(\text{count} | a_{lmys}, \text{NSZ}) = \exp(\beta_0 + \beta_{\text{Density2}} + a_{lmys}) \quad (4.1)$$

$$a_{lmys} \sim N(0, \sigma_{a_{lmys}}^2) \quad (4.2)$$

$$\sigma^2 = \text{var}(\text{count} | a_{lmys}, \text{NSZ}) = \mu(1 + \frac{\mu}{\theta}) \quad (4.3)$$

$$\text{Logit}(p) = \beta_0^{z_i} + \beta_{\text{year}}^{z_i} \quad (4.4)$$

$$\log(\theta) = \beta_0^{(\text{disp})} + \beta_{\text{season}}^{(\text{disp})} \cdot \text{Season} \quad (4.5)$$

Where a_{lmys} are Location, Month, Year and Season specific random effects, NSZ is the event “non-structure zero”, $p = 1 - p_r(NSZ)$ is the zero-inflation probability, and β ’s are regression coefficient with the subscript denoting the covariate/ level (with 0 denoting intercept).

Table 4-2a: Summary of the ZINB-GLMM for the Number of Property crimes

	Parameters	Est.	SE	RR	95% CI	p-value
Random effect	$\sigma_{Location}$	1.38				
	σ_{Month}	0.05				
	σ_{Year}	1.05				
	σ_{Season}	0.10				
Conditional model	Intercept	2.91	0.30	18.31	17.72; 18.90	***
	Density2	-0.79	0.35	0.45	-0.23; 1.13	*
Zero-Inflation model	Intercept	-0.50	0.36	0.60	-0.11; 1.31	***
	Ref (2011)					
	2012	-17.48	2387.26	0	-4679.03; 4679.03	
	2013	5.72	0.43	304.9	304.06; 305.74	***
	2014	-17.53	2385.93	0	-4676.42; 4676.42	
	2015	-0.77	0.50	0.46	-0.52; 1.44	
	2016	-17.53	2385.93	0	-4676.42; 4676.42	
Dispersion model	Intercept	1.96	0.10		-0.196; 0.196	***
	Ref(Summer)					
	Autumn	0.17	0.14	1.19	0.92; 1.46	
	Winter	-0.11	0.14	0.9	0.63; 1.17	
	Spring	-0.18	0.14	0.84	0.57; 1.11	

Estimate (Est.), Standard Error (SE), Relative Risk ratio (OR), Reference category (Ref), Significant level (***=0.001; *=0.05)

The model summary can be broken down into five sections. The first section includes the general overview containing a description of the model specification (family, formula, zero inflation, dispersion, data) together with the information criterion (AIC and BIC).

The second section describes the variability of the random effects. In this model, we only had random effects in the conditional model (equation (4.1)). The estimated standard deviations;

$$\sigma_{Location} = 1.38$$

$$\sigma_{Month} = 0.05$$

$$\sigma_{Year} = 1.05$$

$$\sigma_{Season} = 0.10$$

corresponding to $\sqrt{\sigma_{\alpha_{lmys}}^2}$ in equation (4.2). This indicates how much of the variation in the study of Property crime can be attributed to each random term. The variability of the Number of Property crime according to Month and Season was smaller when compared to that of the Location and Year. This was caused by the high dispersion parameter computed under the two variables. The smaller variation in months and seasons, an indication of less time, is always expected between Property crime cases. To support the view, on the monthly and seasonal perspective, the study has shown that Property crime happens frequently.

The third section describes the relative risk ratio of the conditional model $(\beta_0, \beta_{Density2})$ including a 95% confidence interval and p-values. For the confidence interval, the null value is one since it is estimated on a natural scale. In most cases where a 95% confidence interval does not include the null value the findings are statistically significant. Alternatively, parameters that are statistically significant in the model have a p-value below 0.05 as shown in **Table 4-2a**. Both the intercept $(\beta_0 = 18.31)$ and the relative risk ratio $\beta_{Density2} = 0.45$ are statistically significant. Based on this, it means without considering the population density, approximately 18 Property crimes can be anticipated in Windhoek per annum. The expected counts are conditional on every other value being held constant. That is, including the random Location, Month, Year, and Season effects, population density is expected to have a 45 percent increase on Property crime. In other words, considering the population density, the Number Property crime will increase by 45 percentage.

The fourth section describes the Zero-Inflation model which is like the conditional model except that this model has a logit link. The estimates in this section correspond to β_0^{zi} and β_{year}^{zi} relative risk ratio from equation (4.4). The Zero-Inflation model estimates the probability of an extra zero. The baseline odds of no Property crime reported in Windhoek is 0.6. In addition to that, $\beta_{2013} > 0$ means higher abundances in year 2013 unaffected by population density. This essentially means that during year 2013, the number of Property crimes that were not recorded (but there was an intention) was not due to the number of residents per square kilometre in the area. In contrast, the exploratory study has proven that an area with a high population density experienced high Number of Property crimes during each season.

The confidence interval were estimated using the expression:

$$e^{\widehat{\beta}_i \pm Z_{\frac{\alpha}{2}}(S.E(\widehat{\beta}_i))} \quad (4.6)$$

In this case $\widehat{\beta}_i$ represents the model parameters estimated, S.E is the standard error for the corresponding parameter and Z corresponds to the critical value associated with a 95% degree of confidence. The second component of the model $[Z_{\frac{\alpha}{2}}(S.E(\widehat{\beta}_i))]$ is called the margin of error.

Since $\widehat{\beta}_i = \log(mean) = \log(\mu)$,

hence $e^{\widehat{\beta}_i} = (mean) = \mu$

From **Table 4-2a**, it is obtained that the confidence interval for the estimate $\widehat{\beta}_{2013}$ on the conditional model is 304.06 ; 305.74. It was concluded that this confidence interval provided the study with plausible values for the parameter. If repeated samples were taken and the 95% confidence interval computed for each sample, 95% of the intervals would contain that population parameter. The R-syntax used in setting up the ZINB-GLMM with Number of Property crimes as a response variable is shown in appendix A.

4.5.2 Zero-Inflated Negative Binomial Generalized Linear Mixed Model (ZINB-GLMM) for the Number Violent crime.

For this GLMM modelling, the response variable of interest was Violent crime while the independent variables were *Month*, *Season*, *Location*, *Year* and *Density2*. The first four variables (*Month*, *Season*, *Location* and *Year*) were the random effects. These effects were chosen to be “random” because the crime committed in one month is haphazard of the crime committed in the next month. This is also applicable to *Location*, *Season* and *Year*. However *Density* was chosen to be the fixed effect because the number of people per square kilometer could be measured during this study period. Moreover, the random effects were nested in this study, because each police officer recorded a certain number of cases, and no two officers recorded the same case.

The summary of the ZINB-GLMM from Generalized Linear Mixed Model Template Model Builder (glmmTMB) is modelled on the familiar output format of lme4. The model fitted is:

$$\begin{aligned} \text{Mixedglmm2} < -\text{glmmTMB}(\text{VIOLENT} \sim \text{Density2} + (1|\text{Month}) + (1|\text{Year}) \\ &+ (1|\text{Location}) + (1|\text{Season}), \text{zi} = \sim \text{Year} \\ &+ \text{disp} = \sim \text{Season}, \text{data4}, \text{family} = \text{nbinom2}) \end{aligned} \quad (4b)$$

Following the arguments in the function (Mixedglmm2), this model allows the conditional mean to depend on *Density* and vary randomly by *Location*, *Month*, *Year* and *Season*. It further allows the number of structural (extra) zeros to depend on years. Additionally, it allows the dispersion parameter to depend on the *Season* of the year. This model can be represented by the equations (4.1), (4.2), (4.3), (4.4) and (4.5).

Table 4-2b: Summary of the ZINB-GLMM for the Number of Violent crimes

	Variables	Estimates(β_i)	Standard error	RR	95% CI	p-value
Random effect	$\sigma_{Location}$	1.26				
	σ_{Month}	0.10				
	σ_{Year}	0.64				
	σ_{Season}	0.14				
Conditional model	Intercept	1.62	0.37	5.05309	4.33;5.78	***
	Density2	0.01	0.32	1.01005	0.38;1.64	
Zero-Inflation model	Intercept	-2.94	0.31	0.052866	0.55;0.66	***
	Ref (2011)					
	2012	-18.27	2672.63	1.16E-08	-5238.35;5238.36	
	2013	-18.27	2818.91	1.16E-08	-5525.06;5525.06	
	2014	-18.28	2607.59	1.15E-08	-510.88;5110.88	
	2015	-0.53	0.42	0.588605	-0.23;1.41	
	2016	-16.36	1609.31	7.85E-08	-3154.25;3124.25	
Dispersion model	Intercept	1.56	0.1	4.758821	4.56;4.95	***
	Ref (Summer)					
	Autumn	0.7	0.18	2.013753	1.66;2.37	***
	Winter	0.55	0.18	1.733253	1.38;2.09	**
	Spring	0.1	0.16	1.105171	0.79;1.42	

Estimate (Est.), Standard Error (SE), Relative Risk (RR), Reference category (Ref), Significance level (***=0.001; *=0.05)

The model summary can be broken down into five sections. The first section includes the general overview containing a description of the model specification (family, formula, zero inflation, dispersion, data) together with the information criterion (AIC and BIC).

The second section describes the variability of the random effects. In this model, there were random effects in the conditional model only (equation (4.1)). The estimated standard deviations;

$$\sigma_{Location} = 1.26$$

$$\sigma_{Month} = 0.10$$

$$\sigma_{Year} = 0.64$$

$$\sigma_{Season} = 0.14$$

which are $\sqrt{\sigma_{a_{lmys}}^2}$ in equation (4.2). This indicates how much of the variation in the study of Violent crime can be attributed to each random term. The variability of the Number of Violent crimes due to Month and Season was smaller than that of the Location and Year. This was caused by the high dispersion parameter computed under the two variables. The smaller variations in months and seasons were an indication of less time anticipated between Property crime occurrences in Windhoek.

The third section describes the relative risk ratio of the conditional model $(\beta_0, \beta_{Density})$ including a 95% confidence interval and p-values. The confidence interval is interpreted the same way as that one for ZINB-GLMM for the Number of Property crimes. The intercept $(\beta_0 = 5.05)$ is statistically significant. Based on this result, approximately 5 Violent crimes are expected in Windhoek every year. Considering the population density, will not significantly affect the Number Violent crimes. This may be because most of the Violent crimes are not reported to the police or do not occur in public places as is the case with Property crimes, hence, the number of people in each area will not play a role.

The fourth section describes the Zero-Inflation model which was similar to the conditional model except that this model has a logit link. The estimates in this section correspond to β_0^{zi} and β_{year}^{zi} relative risk ratios from equation (4.4). The Zero-Inflation model estimates the

probability of an extra zero (structural). The baseline odds of no Violent crime reported in Windhoek in a year is 0.05 and it is not significant. The exploratory study has proven that areas with a high population density experienced high Violent crime during each season.

The confidence intervals were estimated using the expression (4.6) and its analogous to that of Property crime. It was observed that none of the estimates in the zero-inflation model were significant.

The last section presents the outcome of the Dispersion model. Based on the confidence intervals as well as the p-values, it was found out that the intercept, Autumn and Winter seasons were significant in the model. The dispersion model accounted for heteroskedasticity. In this study, the Number of Violent crimes were more variable as the year progressed (from April to October or Autumn to winter). This may be due to societal factors that create a climate for violence; factors that reduce inhibitions against violence and those that create tensions within society. These factors may include among others, educational and socio-economic factors that maintain high levels of inequality between groups.

The R-syntax used in setting up the ZINB-GLMM with Violent crime as a response variable is shown in appendix A.

The distribution considered in the **Mixedglmm1** and **Mixedglmm2** is **nbinom2**, returns an overdispersion parameter θ . In contrast to most families, larger θ corresponds to a lower variance which is $\mu(1 + \mu/\theta)$.

Based on the nature of the data, other suggested Property crime models and analogous Violent crime models were fitted, and their AIC values were tabulated. The purpose was to choose a model with a small AIC.

CHAPTER FIVE

5. DISCUSSIONS, CONCLUSIONS AND RECOMMEDATIONS

5.1 DISCUSSIONS

This study has investigated ecological factors influencing the Number of Property and the Number Violent crimes in Windhoek, Namibia. It was found that there are some similarities and disimilarities in literature and the study's findings.

During the study period, an average of 68% of the recorded crime was based on Property while 32% was Violent crime in Windhoek. This result contradicts the statistics by NUMBEO (2018) which indicated that Property and Violent crimes stood at 74.06 and 72.73%, respectively, during the same period. However, both results indicated that the Number Property crimes were higher than the Number of Violent crimes.

Specifically, the Number of Property and the Number of Violent crimes were slightly high during Spring and Winter time. During Spring (October-December) crime increased because numerous residents normally travel for holidays with their families leaving their houses unattended or with no security. This invokes the offender to commit Property crime such as house breaking. Whereas in Winter (July-September), it was assumed that criminals take advantage of the windy and cold weather to commit Property crime as most people prefer to stay indoors due to the cold weather.

However, significant to note is that correlation does not mean causation. A possible explanation is not that there are more people on the streets committing crimes during the holidays and Winter times, but also that the possibility of arresting the offenders is lessened by the fact that there could be less cops patrolling in the streets at that time. Also, it is possible that the crime rate during these times was higher than what the data shows because of a reporting bias. In

general, the Number of Property crimes has slightly increased compared to the Number of Violent crimes between 2011 and 2016.

It was found that there was a direct relationship between the two types of crimes and population densities in Windhoek. This finding seems to bolster the study result by (Hanley et al., 2016) which indicated that population density slightly better in prediction of crimes. The current study results show that most areas with a high population density have a higher crime intensity (with an exemption to Windhoek Central).

Furthermore, those locations with high population densities were considered as the over clouded areas. Residents in these areas often move around leaving their houses unguarded and this attracts criminals. This corresponds to the findings by (Cohen & Felson, 2016) that crimes result from the convergence of some elements such as, suitable target, motivated offender, and the absence of capable guardians. Even though Windhoek Central has a small population density more crime were recorded there as people gather in the area for employment, school and shopping purposes. Using the same logic, affluent areas attract more criminals for theft and robberies due to the opportunities available to them (Justus & Kassouf, 2013). Besides that, the study result strengthen findings that local crime rate are influenced by income deprivation and housing tenure structures (Livingston et al., 2014).

Although Property and Violent crimes were found to be zero inflated, regardless of population density, 18 counts of Property crimes and 5 counts of Violent crimes were expected every year. These data revealed that the Number of Violent crimes have been decreasing while the Number of Property crimes have been constantly increasing as from 2011 to 2016. The zeros (non-structural zeros) obtained were results of no crime recorded within some months or seasons of the study period due to effective policing or lack of suitable targets and this contributed to the choice of ZINB-GLMM. The study indicated that crime data can be modelled using ZINB-

GLMM as an alternative to spatial temporal patterns that were used by other researchers. Perhaps, an important factor to note, is that several other variables were not significant in the model and were not interpreted. However, this is not to say that insignificant findings signal no impact of the indicator on Property or Violent crimes, since individual level driving factors cannot be investigated using police data. More complex models that include more variables influencing the Number of Violent and the Number of Property crimes and interaction among these variables at different levels of aggregation would be preferable.

5.2 CONCLUSION

The criminal activities lie dormant on a daily basis in Windhoek. Crime, no matter how small, has an enormous negative impact on the people's lives. It is basically aimed at terrorising and destroying the nourishing society. Namibia is one of the countries with a high crime rate in the world and needs to further understand the complexity of crime for it to be minimized. The current research is neither an attempt to blame the crime perpetrators or victims of crime, nor a protest for the change of policing strategies. Instead, it advocates for the investigation of the ecological factors influencing the Number of Property and the Number of Violent crimes in Windhoek to provide information and insights leading to more effective and improved prevention strategies of these crimes.

Premised on this background, the study adopted an application of the ZINB-GLMM approach, which evaluates uncertainty in the random effects contributing to the variation in the Number of Property and Number of Violent crimes. The random effects evaluated were based on Month, Year, Season and Location. The results indicate that most of the variation in the study of Property and Violent crimes was due to Location while the effect of Month, Year and Season was not highly significant. Attempting to fit the GLMM model via a GlmmTMB package with non-scaled density variable led to a convergence error.

Furthermore, the study induces a direct relationship between the population Density and the Number of Property and Number of Violent crimes. Higher population Density diminishes the rate of return of legal activities and is more likely to increase the return of illegal activities. Hence, population Density is one of the major contributing factors of the high crime rate in Windhoek. Over clouded areas attract more criminals for theft and robberies due to the opportunities available to them. In addition, there was a strong significant relationship between both types of crimes and the Season of the year.

On average more Property crimes were committed in Windhoek as compared to Violent crimes during Spring and Winter time. Of the crimes that were reported in Windhoek, 68 percent was for Property crimes, while 32 percent was for Violent crimes. It was found that Violent crime happened more often in Goreangab, Independence Arena, Okuryangava and Windhoek Central. In addition to that, Property crime was high in Okuryangava and Windhoek Central. Beside Windhoek Central, these locations were among locations with high population densities. Even though Hakahana has the highest population density, it was found that the number of reported crimes were quite few. Last but not least, one may conclude that there's need to induce a significant relationship between crime and months of the year.

5.3 RECOMMENDATIONS

- 1.) The study recommends that the variance component of ZINB-GLMM be displayed with its standard error to enable researchers to measure the precision of the estimate of the variance components. That is, the smaller the standard error, the more precise the estimate.
- 2.) The study recommends more effective policing in Windhoek during Spring and Winter time, specifically in the areas with high population densities.

- 3.) Windhoek community members should team up to avoid highly diverse population, with little of the social “glue” that binds communities together, in order to effectively reduce crime.
- 4.) Community members should evade practising high levels of residential mobility (where people do not stay for a long time in a dwelling unit) as this motivates the offender to commit crime.
- 5.) Future researchers are advised to use exposure variables to indicate the number of times the crime events could have happened.
- 6.) It is recommended that the Windhoek police should consider the employment status, level of education, age and tribe of the criminal when recording crime to improve data quality in future.
- 7.) It would be advisable for the Windhoek police to geo-code crime data by locations so that future researchers will analyse the spatial aspect of crimes.
- 8.) Future researchers may analyse crime data using spatial temporal patterns and may consider using shapefiles to map crime hotspot.
- 9.) The government should play its role in crime prevention by providing leadership, coordination, and adequate funding and resources. This role can be played, by national, regional or local governments through establishment of:
 - Crime prevention plan with clear priorities and goals
 - Sustainability and accountability of programmes.
- 10.) Those involved in protection services should focus on crime prevention, raising awareness and impose severe punishment to the offender.

References

- Abrahams, N., Mathews, S., Martin, L. J., Lombard, C., Nannan, N., & Jewkes, R. (2016). Gender differences in homicide of neonates, infants, and children under 5 y in South Africa: Results from the cross-sectional 2009 National Child Homicide Study. *PLoS medicine*, 13(4), e1002003.
- Asongu, S., & Kodila-Tedika, O. (2016). Fighting African conflicts and crimes: which governance tools matter? *International Journal of Social Economics*, 43(5), 466-485.
- Ayling, J. (2013). What sustains wildlife crime? Rhino horn trading and the resilience of criminal networks. *Journal of International Wildlife Law & Policy*, 16(1), 57-80.
- Bolker, B. M. (2018). Getting started with the glmmTMB package.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, 24(3), 127-135.
- Boyo, T. M. (2013). *The African brain drain and the social impact of skilled migration*.
- Breetzke, G. D., & Pearson, A. L. (2014). The fear factor: Examining the spatial variability of recorded crime on the fear of crime. *Applied Geography*, 46, 45-52.
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017a). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R journal*, 9(2), 378-400.
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017b). Modeling zero-inflated count data with glmmTMB. *bioRxiv*, 132753.
- Cameron, A. C., & Trivedi, P. K. (2013). *Regression analysis of count data* (Vol. 53): Cambridge university press.

- Chamberlain, A. W., & Hipp, J. R. (2015). It's all relative: Concentrated disadvantage within and across neighborhoods and communities, and the consequences for neighborhood crime. *Journal of Criminal Justice*, 43(6), 431-443.
- Clark, T. S., & Linzer, D. A. (2015). Should I use fixed or random effects? *Political Science Research and Methods*, 3(2), 399-408.
- Cohen, L. E., & Felson, M. (2016). Social Change and Crime Rate Trends: A Routine Activity Approach (1979). In *Classics in Environmental Criminology* (pp. 203-232): CRC Press.
- Council, O. S. A. (2010). South Africa 2010 Crime & Safety Report. *United States Department of State Bureau of Diplomatic Security*.
- de Melo, S. N., Matias, L. F., & Andresen, M. A. (2015). Crime concentrations and similarities in spatial crime patterns in a Brazilian context. *Applied Geography*, 62, 314-324.
- Dore, P. (2013). Urbanisation, Inequality and Youth. *State of the Urban Youth, India 2012*, 75.
- Dunn, P. K., & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3), 236-244.
- Ellis, S., & Shaw, M. (2015). Does organized crime exist in Africa? *African Affairs*, 114(457), 505-528.
- Eze, I. R.-A. (2016). *Exploring passion killing and its implications on the academic wellbeing of university students in Botswana and Namibia*. University of the Western Cape,
- Flores-Agreda, D., & Cantoni, E. (2019). Bootstrap estimation of uncertainty in prediction for generalized linear mixed models. *Computational Statistics & Data Analysis*, 130, 1-17.

- Fournier, D. A., Skaug, H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., . . . Sibert, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, 27(2), 233-249.
- Garcia-Moreno, C., Jansen, H. A., Ellsberg, M., Heise, L., & Watts, C. H. (2006). Prevalence of intimate partner violence: findings from the WHO multi-country study on women's health and domestic violence. *The lancet*, 368(9543), 1260-1269.
- Gelman, A., & Hill, J. (2014). *Data analysis using regression and multilevel hierarchical models* (Vol. 1): Cambridge University Press New York, NY, USA.
- Green, S. B., & Salkind, N. J. (2016). *Using SPSS for Windows and Macintosh, Books a la Carte*: Pearson.
- Han, L., Bandyopadhyay, S., & Bhattacharya, S. (2013). Determinants of violent and property crimes in England and Wales: A panel data analysis. *Applied Economics*, 45(34), 4820-4830.
- Hanley, Q. S., Lewis, D., & Ribeiro, H. V. (2016). Rural to urban population density scaling of crime and property transactions in English and Welsh parliamentary constituencies. *PloS one*, 11(2), e0149546.
- Harrison, X. A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ*, 2, e616.
- Janko, Z., & Popli, G. (2015). Examining the link between crime and unemployment: a time-series analysis for Canada. *Applied Economics*, 47(37), 4007-4019.
- Jones, R. W., & Pridemore, W. A. (2012). The Foreclosure Crisis and Crime: Is Housing-Mortgage Stress Associated with Violent and Property Crime in US Metropolitan Areas? *Social Science Quarterly*, 93(3), 671-691.

- Justus, M., & Kassouf, A. L. (2013). Evidence of the effect of wealth observed by criminals on the risk of becoming a victim of property crimes. *Economía*, 14(2), 88-101.
- Kemper, J., & ROUX, J. P. (2005). Of squeezers and skippers: factors determining the age at moult of immature African Penguins *Spheniscus demersus* in Namibia. *Ibis*, 147(2), 346-352.
- Khan, N., Ahmed, J., Nawaz, M., & Zaman, K. (2015). The socio-economic determinants of crime in Pakistan: New evidence on an old debate. *Arab Economic and Business Journal*, 10(2), 73-81.
- Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *The lancet*, 360(9339), 1083-1088.
- Linning, S. J., Andresen, M. A., & Brantingham, P. J. (2017). Crime seasonality: Examining the temporal fluctuations of property crime in cities with varying climates. *International journal of offender therapy and comparative criminology*, 61(16), 1866-1891.
- Livingston, M., Kearns, A., & Bannister, J. (2014). Neighbourhood structures and crime: the influence of tenure mix and other structural factors upon local crime rates. *Housing Studies*, 29(1), 1-25.
- Magnusson, A., Skaug, H., Nielsen, A., Berg, C., Kristensen, K., Maechler, M., . . . Brooks, M. M. (2017). Package 'glmmTMB'.
- Mazzitelli, A. L. (2007). Transnational organized crime in West Africa: the additional challenge. *International Affairs*, 83(6), 1071-1090.
- McGrory, S. (2018). The influence social problems have with violent crime and its impact on society: An investigation on the five boroughs of New York. In.
- McShane, M. D., & Williams, F. P. (1997). *Victims of Crime and the Victimization Process* (Vol. 6): Taylor & Francis.

- Moffatt, S., & Goh, D. (2013). An update of long-term trends in property and violent crime in New South Wales: 1990-2012. *Crime and Justice Statistics Bureau Brief. Issue(84)*.
- Neema, I., & Böhning, D. (2012). Monitoring murder crime in Namibia using Bayesian space-time models. *Journal of Probability and Statistics*, 2012.
- NSA. (2011). *Namibia 2011: Population & Housing Census Main Report*: Namibia Statistics Agency.
- NUMBEO. (2018, January 2018). Crime in Windhoek, Namibia. Retrieved from <https://www.numbeo.com/crime/in/Windhoek>
- Palmary, I., Rauch, J., & Simpson, G. (2014). Violent crime in Johannesburg. In *Emerging Johannesburg* (pp. 117-138): Routledge.
- Papps, K., & Winkelmann, R. (2000). Unemployment and crime: New evidence for an old question. *New Zealand Economic Papers*, 34(1), 53-71.
- Posada, D., & Buckley, T. R. (2004). Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology*, 53(5), 793-808.
doi:10.1080/10635150490522304
- Raine, A. (2013). The anatomy of violence: The biological roots of crime.
- Sandema, R. (2005). *Social factors influencing crime-armed robbery-in Windhoek from 1997-2000*. University of Namibia,
- Schepers, D. (2017). Causes of the causes of juvenile delinquency: Social disadvantages in the context of Situational Action Theory. *European journal of criminology*, 14(2), 143-159.
- Schielzeth, H., & Nakagawa, S. (2013). Nested by design: model fitting and interpretation in a mixed model era. *Methods in Ecology and Evolution*, 4(1), 14-24.

- Shaw, M., & Reitano, T. (2013). The evolution of organised crime in Africa-towards a new response. *Institute for Security Studies Papers*, 2013(244), 28.
- Song, W., & Liu, D. (2013). Exploring spatial patterns of property crime risks in Changchun, China. *International Journal of Applied Geospatial Research (IJAGR)*, 4(3), 80-100.
- Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science*, 57(3), 748-761.
- Team, R. (2013). R development core team. *RA Lang Environ Stat Comput*, 55, 275-286.
- Tonry, M. (2014). Why crime rates are falling throughout the Western world. *Crime and justice*, 43(1), 1-63.
- UNODC. (2010). *International statistics on crime and justice*. Retrieved from Helsinki: <https://www.unodc.org/documents/data-and-analysis/Crime->
- Vogel, M., & Van Ham, M. (2018). Unpacking the relationships between impulsivity, neighborhood disadvantage, and adolescent violence: an application of a neighborhood-based group decomposition. *Journal of youth and adolescence*, 47(4), 859-871.
- Walker, S. (2018). *Crisis-Opportunity, Liability-Asset: Governing Vacant Land Reuse in Cleveland, Ohio*.
- Wambua, P. M. (2015). Call the police? Across Africa, citizens point to police and government performance issues on crime.
- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv preprint arXiv:1308.5499*.
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, 27(8), 1-25.

Appendix A

This section outline how the dataset is loaded into R with their syntax.

The data is loaded into R simply as csv file using the syntax,

```
data1<-read.csv(file.choose(),header=T)
```

Table 1A: Initial header line for the study dataset

Year	Location	Month	Seasonn	Season	PROPERTY	VIOLENT	Density	Density2
1	1	1	Summer	1	19	1	1013	0.1013
1	1	2	Summer	1	24	5	1013	0.1013
1	1	3	Summer	1	29	11	1013	0.1013
1	1	4	Autumn	2	33	10	1013	0.1013
1	1	5	Autumn	2	33	11	1013	0.1013
1	1	6	Autumn	2	40	5	1013	0.1013

Four of the explanatory variables (Year, Location, Month and Season) were converted into #factors prior to data presentation.

```
data1$Year<-as.factor(data1$Year)
data1$Location<-as.factor(data1$Location)
data1$Month<-as.factor(data1$Month)
data1$Seasonn<-as.factor(data1$Seasonn)
data1$Season<-as.factor(data1$Season)
> str(data1)
'data.frame': 1969 obs. of 9 variables:
 $ Year : Factor w/ 6 levels "1","2","3","4",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ Location: Factor w/ 59 levels "1","2","3","4",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ Month : Factor w/ 12 levels "1","2","3","4",...: 1 2 3 4 5 6 7 8 9 10 ...
 $ Seasonn : Factor w/ 5 levels "", "autumn", "spring",...: 4 4 4 2 2 2 5 5 5 3 ...
 $ Season : Factor w/ 4 levels "1","2","3","4": 1 1 1 2 2 2 3 3 3 4 ...
 $ PROPERTY: int 19 24 29 33 33 40 35 37 44 33 ...
 $ VIOLENT : int 1 5 11 10 11 5 6 7 8 7 ...
 $ Density : int 1013 1013 1013 1013 1013 1013 1013 1013 1013 1013 ...
 $ Density2: num 0.101 0.101 0.101 0.101 0.101 0.101 ...
```

Table 2A: Coded variables

Variables	Codes
Year	Year
2011	1
2012	2
2013	3
2014	4
2015	5
2016	6
Month	Month
January to December	1 to 12
Season	
Summer (January-March)	1
Autmun (April-June)	2
Winter (July-September),	3
Spring (October-December)	4
Location	Location
Academia	1
Babilon	2
Brake Water	3
Cimbebasia	4
Damara Location	5
Dolam	6
Donkerhoek	7

Dorado Park	8
Dorando Valley	9
Eehambo DaNehale	10
Eros	11
Eros airport	12
Freedom square	13
Freedomland	14
Gemeente	15
Golgota	16
Goreangab	17
Green-well Matongo	18
Grysbloek	19
Hakahana	20
Havana	21
Herero location	22
Hockland park	23
Independence Arena	24
Khomasdal	25
Kilimajaro	26
Kleine Kuppe	27
Klein Windhoek	28
Lafrenz	29
Ludwigsdorf	30
Marua Mall	31
Malaka Draai	32

Maroela	33
Mix	34
Mukwanangombe	35
Mukwanekamba	36
Northern Industrial	37
Nubuamis	38
Okahandja Park	39
Okuryangava	40
Olympia	41
Ombili	42
One-nation	43
Ongulumbashe	44
Oshitenda	45
Otjomuise	46
Pionierspark	47
Properita	48
Rocky Crest	49
Shandumbala	50
Single Quarter	51
Southern Industrial	52
Soweto	53
Suiderhof	54
Vambo location	55
Wanahenda	56
Windhoek Central	57

Windhoek North	58
Widhoek West	59

The R-syntax used in constructing the Property and corresponding Violent crime models are shown below.

Generalized linear mixed model (GLMM)

```
require(glmmTMB) #Loading GLMM package
```

```
Mixedglmm5
```

```
<- glmmTMB(PROPERTY~Density2 + (1|Month) + (1|Year)
+ (1|Location) + (1|Season), zi = ~Year, disp
= ~Season, data1, family = nbinom2) # GLMM for property crime
```

```
> summary(Mixedglmm5) #Display the model output
```

```
Mixedglmm6 <- glmmTMB(VIOLENT~Density2 + (1|Month) + (1|Year)
+ (1|Location) + (1|Season), zi = ~Year, disp
= ~Season, data1, family = nbinom2) # GLMM for Violent crime
```

```
> summary(Mixedglmm6) #Display the model output
```

Poisson Model

```
Property.Pois <- glm(PROPERTY~ Month + Year + Location + Density2, data
+ data1, family = poisson) # Poisson model for Property crime
```

```
summary(Property.Pois) #Display the model output
```

```
VIOLENT.Pois <- glm(VIOLENT~ Month + Year + Location + Density2, data
+ data1, family = poisson) # Poisson model for Violent crime
summary(VIOLENT.Pois) # Display the model output
```

Negative binomial (NB)

```
NB <- glm.nb(PROPERTY~Month + Year + Location + Density2, data
= data1) # NB model for Property crime
```

```
summary(NB) # Display the model output
```

```
NB <- glm.nb(VIOLENT~Month + Year + Location + Density2, data
= data1) # NB model for Violent crime
```

```
summary(NB) # Display the model output
```

Zero-inflated Poisson (ZIP)

Property.ZIP

```
<- zeroinfl(PROPERTY~Month + Year + Location + Density2, data
= data1, link = logit, dist = poisson, trace
= TRUE) # ZIP model for Property crime
```

```
summary(Property.ZIP) # Display the model output
```

```
Violent.ZIP <- zeroinfl(VIOLENT~Month + Year + Location + Density2, data
= data1, link = logit, dist = poisson, trace
= TRUE) # ZIP model for Violent crime
```



```
summary(Violent.ZIP) #Display the model output
```

Zero-inflated negative binomial (ZINB)

```
Property.ZINB
```

```
< -zeroinfl(PROPERTY~Month + Year + Location + Density2, data  
= data1, link = "logit", dist = "negbin", trace = TRUE, EM  
= FALSE) # ZINB model for Property crime
```

```
summary(Property.ZINB) #Display the model output
```

```
Violent.ZINB < -zeroinfl(VIOLENT~Month + Year + Location + Density2, data  
= data1, link = "logit", dist = "negbin", trace = TRUE, EM  
= FALSE) # ZINB model for Violent crime
```

```
summary(Violent.ZINB) #Display the model output
```

ZINB model for Number of Property crime.

			Standard		
	Covariate	Estimates	error	IRR/OR	P-value
Count model (negbin with log link)	Intercept	3.200	0.095	24.533	0.001
	Month	0.013	0.008	1.013	0.1
	Year	-0.002	0.014	0.998	1
	Location	0.018	0.001	1.018	0.001
	Density	-0.244	0.06	0.783	0.001
	Log(theta)	-0.098	0.04	0.907	0.05
Zero-inflation model (logit link)	Intercept	-1.412	0.316	0.244	0.001
	Month	-0.052	0.025	0.949	0.05
	Year	-0.236	0.045	0.790	0.001
	Location	0.012	0.005	1.012	0.05
	Density	-0.147	0.194	0.863	1

ZINB model for Number of Violent crimes.

	Covariate	Estimates	Standard error	IRR/OR	sig.
Count model (negbin with log link)	Intercept	1.833	0.103	6.253	0.001
	Month	0.023	0.008	1.023	0.01
	Year	-0.004	0.016	0.996	1
	Location	0.013	0.002	1.013	0.001
	Density2	0.993	0.072	2.699	0.001
	Log(theta)	-0.253	0.032	0.776	0.001
Zero-inflation model (binom with logit link)	Intercept	7.875	58.178	2630.686	
	Month	-0.220	0.169	0.803	
	Year	-9.792	58.165	0.000	
	Location	-0.036	0.026	0.965	
	Density2	1.169	0.539	3.219	0.05

Appendix B

This section consists of R syntax that were used for plotting figures as well as the materials used for data collection.

```
install.packages("lattice") # installing a package lattice
```

```
install.packages("ggplot2") #installing a package ggplot2
```

```
Library(lattice) # loading lattice package for use in plotting
```

```
Library(ggplot2) # loading ggplot2 package for use in plotting
```

```
#Property and Violent crime distribution for each year
```

```
boxplot(PROPERTY~Year,data = data4,col = "orange",xlab = "years",ylab  
= "Property crime(count)",ylim = c(0,200))
```

```
boxplot(VIOLENT~Year,data = data4,col = "orange",xlab = "years",ylab  
= "Violent crime(count)",ylim = c(0,200))
```

```
#Property and Violent crime distribution for each month
```

```
boxplot(PROPERTY~Month,data = data4,col = "orange",xlab  
= "Months",ylab = "Property crime(count)",ylim = c(0,150))
```

```
boxplot(VIOLENT~Month,data = data4,col = "orange",xlab  
= "Months",ylab = "Violent crime(count)",ylim = c(0,150))
```

```
# Property and Violent crime plot for each season
```



```
boxplot(PROPERTY~Seasonn,data = data4,col = "orange",xlab
= "Seasons",ylab = "Property crime(count)",ylim = c(0,150))
```

```
boxplot(VIOLENT~Seasonn,data = data4,col = "orange",xlab
= "Seasons",ylab = "Violent crime(count)",ylim = c(0,150))
```

#Property and Violent crime plot for each season per population density

```
xyplot(PROPERTY~Density2|Seasonn,data4,index
= function(x,y)coef(lm(y~x))[1],xlab
= "scaled population Density",ylab
= "AVERAGE PROPERTY CRIME",aspect = "xy")
```

```
xyplot(VIOLENT~Density2|Seasonn,data4,index
= function(x,y)coef(lm(y~x))[1],xlab
= "scaled population Density",ylab
= "AVERAGE VIOLENT CRIME",aspect = "xy")
```

```
ggplot(data1,aes(x = Location,y = VIOLENT)) + geom_point() #Location vs Violent
crime plot
```

```
ggplot(data1,aes(x = Location,y
= PROPERTY)) + geom_point()#Location vs Property crime
```

Specifying the distribution for Property and Violent crime

```
library(car)# loading the car package
```

```
library(MASS)# loading the MASS package
```

```
#Normality
```

```
data4$PROPERTY.t < -data4$PROPERTY + 1
```

```
qqp(data4$PROPERTY.t,"norm")
```

```
data4$VIOLENT.t < -data1$VIOLENT + 1
```

```
qqp(data4$VIOLENT.t,"norm")
```

```
attach(data1) # attaching the data set
```

```
qqnorm(data1$PROPERTY) # Normal Q - Q plot for PROPERTY CRIME
```

```
qqline(data1$PROPERTY,col
```

```
= 2) #Attach a normal Q - Q line for PROPERTY CRIME
```

```
shapiro.test(data1$PROPERTY)# Perform a normality test
```

```
qqnorm(data1$PROPERTY) # Normal Q - Q plot for VIOLENT CRIME
```

```
qqline(data1$VIOLENT,col = 2) #Attach a normal Q - Q line for VIOLENT CRIME
```

```
shapiro.test(data1$VIOLENT)# Perform a normality test
```


Negative Binomial (NB)

```
#loading neccessary packages for accessing negative binomial distribution
```

```
require(car)
```

```
require(MASS)
```

```
#NB
```

```
nbinom <- fitdistr(data4$PROPERTY.t, "Negative Binomial")
```

```
qqp(data4$PROPERTY.t, "nbinom", size = nbinom$estimate[[1]], mu  
    = nbinom$estimate[[2]])
```

```
nbinom <- fitdistr(data4$VIOLENT.t, "Negative Binomial")
```

```
qqp(data4$VIOLENT.t, "nbinom", size = nbinom$estimate[[1]], mu  
    = nbinom$estimate[[2]])
```

```
#Long normal
```

```
qqp(data4$PROPERTY.t, "lnorm")
```

```
qqp(data4$VIOLENT.t, "lnorm")
```

Best fit model

Syntax for assessing the best fit model

```
Library (DHARMA)                # support for glmmTMB through simulation  
install.packages("gap")          # in support of residual plot  
simulationOutput <- simulateResiduals(fittedModel = Mixedglmm5, n = 250)  
simulationOutput$scaledResiduals # display simulated values  
plot(simulationOutput)           # display residual plot
```

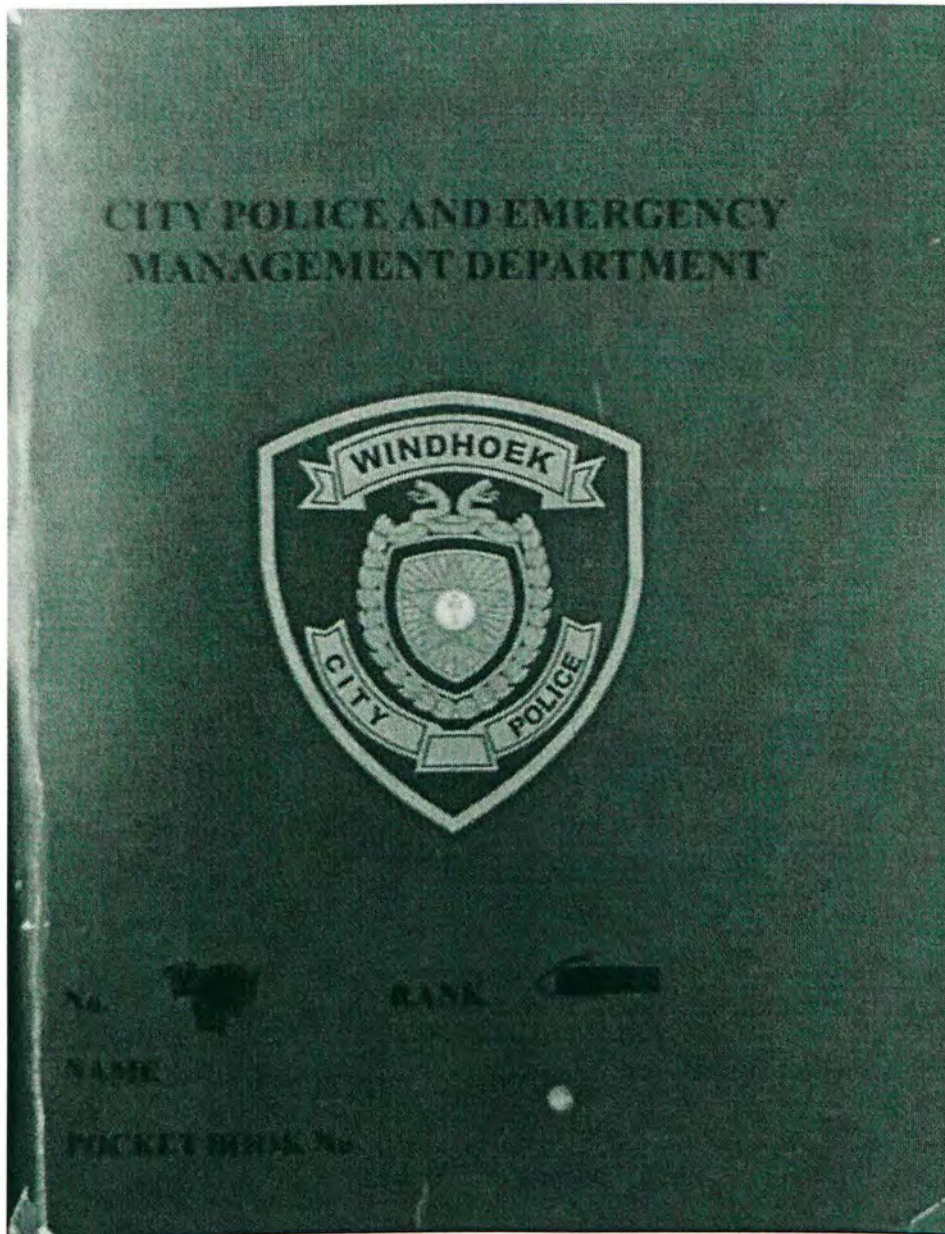
```
pearsonres1 <- residuals(Violent.zinb, type = pearson) # Pearson residual for Violent.zinb  
plot(predict(Violent.zinb), pearsonres, xlab = "Fitted values", ylab =  
"Pearson Residuals", ylim = max(abs(pearsonres1)) * c(-1,1)) #Pearson residual plot
```

```
pearsonres2 <- residuals(Property.zinb, type = pearson) # Pearson residual for  
property.ZINB  
plot(predict(Property.zinb), pearsonres, xlab = "Fitted values", ylab =  
"Pearson Residuals", ylim = max(abs(pearsonres2)) * c(-1,1)) #Pearson residual plot
```

>

Pocket book

Pocket book is a book used by any member of city police who attend to a certain complaint or crime. It is used in recording information pertaining to crime.



Letter addressed to city police Chief A. K. Kanime



Faculty of Science
Department of Statistics and Population Studies

To
Chief A. K. Kamine
City Police Service
City of Windhoek

24th January, 2018

Dear Sir,

R: Request for the Annual Crime Report

I trust this letter will find you in good health.

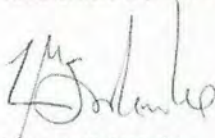
I write to introduce Mr Jonas Amunyela, a Masters student in the Department of Statistics and Population Studies, here at UNAM.

Mr Amunyela is generating a research proposal which will require analysis of crime statistics. Considering that the City Police Service compiles reports on crime in the City, I think it will be relevant to have such reports to enhance the background and provide a strong justification to the research topic he intends to pursue.

I therefore write for your kind consideration to assist with copies of crime reports for the previous years. Such reports will assist Mr Amunyela to generate a plausible background in support of his intended research.

I trust this is in order.

I look forward to your usual assistance.


Lawrence N. Kazembe
Head of Department

