ZERO-AUGMENTED MODELS FOR EXPLORING THE FACTORS AFFECTING THE PASS RATE OF GRADE 10 LEARNERS IN KHOMAS REGION IN 2016

A 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

RAPIKAMA MUMBUU

9231684

APRIL 2020

MAIN SUPERVISOR: PROF. L. PAZVAKAWAMBWA
(UNAM, DEPARTMENT OF STATISTICS AND POPULATION STUDIES)

CO-SUPERVISOR: DR. O. OYEDELE
(UNAM, DEPARTMENT OF STATISTICS AND POPULATION STUDIES)
ABSTRACT

The poor performance of grade 10 learners has been a big concern over the last few years and in the effort to understand this phenomenon there has been efforts to present models that explain it. Modelling semi-continuous data with the presence of excess zeros has become a common phenomenon in real life situations. Common models such as linear models cannot handle zero-inflated data.

This study aimed at exploring the factors which influence Khomas Region grade 10 learners’ pass rate using Generalized Linear Models (GLM). The data used for this study was obtained from the Directorate of National Examination and Assessment for the year 2016, with permission from the Permanent Secretary of the Ministry of Education. Descriptive statistics were used to describe the socio-demographic variables, namely: age, location and type of school. Six GLMs were fitted (Poisson, Negative Binomial, Hurdle Poisson, Hurdle Negative Binomial, Zero Inflated Poisson and Zero- Inflated Negative Binomial) to assess their goodness of fit on modelling the zero-inflated DNEA count data. The goodness of fit of each model was determined using the Akaike Information Criterion (AIC) value. All analyses were done using the R software version 3.3.1, with its MASS, pscl, and AER packages, as well as the Statistical Package (SPSS).

The Zero- Inflated Negative Binomial performed better based on its lowest AIC values among the six fitted GLMs. The results revealed that the age of the learner, school location and the type of school (private/state) had significant differential in pass rate with p-values less than 0.05 in the Zero- Inflated Negative Binomial model.
In densely populated areas, emphasis should be put on building more schools in these areas so that classrooms are not overcrowded per subject. In addition, overaged learners should also be given extra assistance such as extra classes and extra motivation. Government should equip state schools to enable incorporation of visuals and other resources stimulate the learners’ interest.

**Keywords**: Semi-continuous data, zero augmented models, pass rate, Namibia
Table of Contents

ABSTRACT ................................................................................................................................. i

List of Tables ............................................................................................................................ vi

List of Figures .......................................................................................................................... vii

Abbreviations/ Acronyms ........................................................................................................ viii

Acknowledgements ................................................................................................................ x

Dedication ................................................................................................................................. xi

CHAPTER 1: INTRODUCTION ................................................................................................. 1

1.1 Orientation of the study ...................................................................................................... 1

1.2 Problem Statement ........................................................................................................... 5

1.3 Objectives of the study ...................................................................................................... 6

1.4 Hypothesis of the study .................................................................................................... 7

1.5 Significance of the study .................................................................................................. 7

1.6 Limitations of the study ................................................................................................... 8

1.7 Delimitations of the study ............................................................................................... 8

1.8 Organisation of the study ................................................................................................. 9

CHAPTER 2: LITERATURE REVIEW ..................................................................................... 10

2.1 Introduction ....................................................................................................................... 10

2.2 Theoretical approaches of academic performance .......................................................... 10

2.3 Factors leading to poor performance in secondary schools .......................................... 11

2.4 Performance of learners in secondary school .................................................................. 17

2.5 Gaps identified ................................................................................................................ 23

2.6 Statistical models for count data .................................................................................... 24

2.6.1 The hurdle model ....................................................................................................... 24

2.6.1.1 The Poisson hurdle model .................................................................................... 28
2.6.1.2. Hurdle negative binomial regression ................................................................. 30
2.6.2 Zero-inflated count models .................................................................................... 30
2.6.2.1 A zero-inflated negative binomial regression model (With hidden Markov chain) 31
2.6.3 Hurdle model versus ZIP model ............................................................................ 32
2.7 Application of the hurdle models and Zero inflated models .................................. 33
2.7.1 The hurdle models .................................................................................................. 33
2.7.2 The Zero inflated models ....................................................................................... 35
2.8 Conceptual framework ............................................................................................... 35

CHAPTER 3: METHODOLOGY ......................................................................................... 37
3.1 Introduction ................................................................................................................. 37
3.2 Research design ......................................................................................................... 37
3.3 Population and samples ............................................................................................ 37
3.4 Variables .................................................................................................................... 37
3.5 Methods of analysis .................................................................................................. 38
3.6 Data collection procedures ....................................................................................... 39
3.7 Ethical considerations ............................................................................................... 39

CHAPTER 4: RESULTS AND DISCUSSIONS .................................................................. 40
4.1 Introduction ................................................................................................................. 40
4.2 Part I: Descriptive statistics ..................................................................................... 40
4.2.1 Profile of the Pass Rate across Sex ...................................................................... 40
4.2.3 Profile of the types of location .............................................................................. 42
4.2.4 Profile of schools and location in Khomas region ............................................... 43
4.3 Part II: Inferential statistics ..................................................................................... 44
4.3.1 Zero-Inflated Negative Binomial model ............................................................... 46
4.4 Discussions ............................................................................................................... 48
4.4.1 Age ....................................................................................................................... 49
4.4.2  Sex ........................................................................................................... 49
4.4.3  School environment .................................................................................. 50
4.4.4  School type ............................................................................................... 51
4.5   Conclusion ................................................................................................... 51

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS .......................................... 53
5.1   Introduction .................................................................................................. 53
5.2   Conclusions ................................................................................................. 53
5.3   Recommendations ....................................................................................... 54

References ............................................................................................................. 56

Appendix A: R-codes ............................................................................................ 60
Appendix B: Models Results .................................................................................. 63
Appendix C: AIC results ....................................................................................... 70
List of Tables

Table 1  Variables summary  
Table 2  Pass rate vs Sex  
Table 3  Distribution of the types of school across school location  
Table 4  Cross tabulation of names of school and location in Khomas region  
Table 5  Various GLMs for modelling count data and their AIC’S and log-likelihood  
Table 6  Zero-Inflated Negative Binomial model coefficients (Binomial with logit link)  
Table 7  Zero-Inflated Negative Binomial model coefficients (with log link)
List of Figures

Figure 1  Variables Affecting Learner Academic Performance  36
Figure 2  Bar chart showing the distribution of the number of subjects passed.  41
Figure 3  Grade 3 learners of the attached primary school at Okayama city.  55
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian Information Criterion</td>
</tr>
<tr>
<td>DNEA</td>
<td>Directorate of National Examination and Assessment</td>
</tr>
<tr>
<td>GLM</td>
<td>Generalized Linear Model</td>
</tr>
<tr>
<td>HAD</td>
<td>High Density Area</td>
</tr>
<tr>
<td>JSC</td>
<td>Junior Secondary Certificate</td>
</tr>
<tr>
<td>LDA</td>
<td>Low Density Area</td>
</tr>
<tr>
<td>NIED</td>
<td>Namibia Institute for Educational Development</td>
</tr>
<tr>
<td>NJSE</td>
<td>Namibia Junior Secondary Examinations</td>
</tr>
<tr>
<td>NPML</td>
<td>Nonparametric Maximum Likelihood</td>
</tr>
<tr>
<td>NSAT</td>
<td>National Secondary Assessment Test</td>
</tr>
<tr>
<td>PQL</td>
<td>Penalised Quasi-Likelihood</td>
</tr>
<tr>
<td>SACMEQ</td>
<td>Southern African Consortium for Monitoring Educational Quality</td>
</tr>
<tr>
<td>SEB</td>
<td>Socio-Economic Background</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package for Social Sciences</td>
</tr>
<tr>
<td>SES</td>
<td>Socio-Economic Status</td>
</tr>
<tr>
<td>ZINB</td>
<td>Zero-Inflated Negative Binomial</td>
</tr>
<tr>
<td>ZIP</td>
<td>Zero Inflated Poisson</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to acknowledge my supervisors, Prof. L. Pazvakawambwa and Dr. O. Oyedele for their dedicated mentoring skills and patience to see me through this huge amount of research work. If it was not for their guidance, advice and assistance the project could not have come to completion. I would also like to acknowledge the assistance I got from the head of department, Professor L Kazembe from UNAM and from colleagues at the Namibia University of Science and Technology (NUST).
Dedication

It is my pleasure to dedicate this work entirely to my wife for her huge and everlasting love, encouragement and care she has provided me with and to my children for their understanding when I was not available to play with them. I am greatly indebted to all the people who made this research work a success, especially for their time and advices.
Declaration

I, Rapikama Mumbuu, 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 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, Rapikama Mumbuu, 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.

.......................... .......................... .......................... ..........................
Name of student         Signature             Date
CHAPTER 1: INTRODUCTION

1.1 Orientation of the study

In the era of globalization and technological revolution, education is considered a first step for every human activity (Saxton, 2008). According to Saxton (2008), education plays a vital role in the development of human capital and is linked to an individual’s well-being and opportunities for a better living. Education is one of the most powerful instruments known for reducing poverty and for laying the basis for a sustainable economic growth. It ensures the acquisition of knowledge and skills that enable individuals to increase their productivity and improve their quality of life (Battle & Lewis, 2002). This increase in productivity also results in more employment opportunities which enhance the economic growth of a country. In addition, education can be viewed as a process through which the intellectual, moral capacities, proper conduct, and technical competency of individuals are developed to make them cultural members of their respective societies (Tuan, 2009).

Secondary education plays a vital role in ensuring that a country’s development through training is a pre-requisite to economic growth and Social development (World Bank, 2008). For this reason, governments all over the world are committed to the provision of education to their citizens. Kirera, (2013) also confirmed the importance of secondary education. She emphasised that secondary school is very important for many individuals because it improves their standard of living when they get jobs, and determines their future life chances and mobility.
A Study by Miller-Grandvaux and Yoder (2012) on secondary schools education revealed that the main challenges in secondary school education seem to be the academic performance of learners. Generally, the academic performance of learners varies from learner to learner, school to school, location to location and country to country. Anecdotal evidence indicates that the school location, environment, inadequate facilities and infrastructure are some of the factors that account for the differences in academic performance of learners across different subjects (Miller-Grandvaux & Yoder, 2012).

At Secondary school level in Namibia, the National Secondary Assessment Test (NSATs) serves to, among others, help diagnose difficulties of weaker learners for remedial action and to identify areas of strengths and weaknesses in class (Ministry of Education (b), 2008). In general, Grades 5 and 7 learners have consistently been performing poorly in NSATs, particularly in reading comprehension and basic numeracy skills. In addition to the NSATs, the Southern African Consortium for Monitoring Educational Quality (SACMEQ) conducts tests for Grade 6 learners in different Southern African countries, including Namibia, to test learners’ literacy and numeracy skills. In the SACMEQ tests, most of the test items are based on texts which learners have to read and then answer questions. These results have shown that Namibia is struggling with literacy and numeracy skills with results that are below 50% (Ministry of Education , 2010).

In Namibia, the education system is divided into three stages, namely primary level, secondary level and tertiary level. The primary level, Grade 1 to 7, prepares children for secondary education. In other words, primary education is the basic education
provided at primary school level. On the other hand, the secondary level stretches over a period of 5 years from Grade 8 to Grade 12 (Namibia Government, 2001). Learners are presented with a Junior Secondary School Certificate (JSC) after successfully completing Grade 10, and they get a Senior Secondary Certificate at the end of Grade 12 (NSSC).

To pass Grade 10, a learner is expected to score 23 points in six subjects including English. All learners in Grade 10 do English first or second language, another local language, Mathematics, Physical Science, Life Science, Geography, History, Entrepreneurship, and Agriculture/Accounting. The learners may only choose either Accounting or Agriculture to complete the list of nine subjects. The recommended age for grade 10 is 16 years and the grading system convert symbols to points as follows: A=7 , B= 6 , C= 5, D = 4, E = 3, F = 2, G =1 and U = 0 . A good performance is regarded as scoring an average of 6D’s.

Learners can choose to enrol at either private or public (government) schools for their education in Namibia. A private school is a school which is established and maintained at the owner's expense while a public school is owned by the state/government (Namibia Government, 2001). The majority of learners in Namibia opt to go for public school because private schools are unaffordable. Education at primary school level in government schools is free in Namibia hence the choice of public schools. This was an initiative by the Namibian government to ensure that all Namibian children are educated up to secondary level, and to ease the financial burden on their parents/guardians.
According to Pandey and Goval (2009), learners in private schools tend to perform much better than their counterparts in government schools. Several reasons have been given to explain the learners’ poor performance in government schools. These include poor school resources, poverty and the illiteracy of parents/guardians. Recent researches further highlight the pervasiveness of teacher absenteeism and the inactivity in government schools as some of the prime reasons for the poor performance (Obiero, Mwebi, & Nyang’ara, 2017). In addition, there are few consistent differences in infrastructure between private and government schools, and private schools tend to have significantly lower pupil-teacher ratios in a classroom.

In a study conducted by Hilongwa (2011) which focused on factors that inhibit learners’ comprehension of the English language at secondary school level, the focus was not on whether teachers teach learners reading comprehension strategies or not. Rather, it focused on finding out why second language learners of English find it difficult to understand texts written in English. She found that Namibian learners’ difficulty in comprehending such texts was attributed to the fact that they did not have much knowledge of the English language and had a limited range of English vocabularies (Peacock, 2013). In the policy brief of the SACMEQ III report of 2011, it was suggested to the National Institute for Educational Development (NIED) and Regional Directors of Education in Namibia that an investigation be carried out on the reasons for low reading comprehension and numeracy skills among learners (Bruwer, 2013).

Since 1993, grade 10 learners in Namibia, regardless of the type of school attended have written the National Junior Secondary Examinations (NJSE) administered by the
Directorate of National Examinations and Assessment (DNEA) in Namibia. The NJSE is compulsory for all registered grade 10 learners in Namibia and is used to assess the achievement of learners in a curriculum in order to provide an estimate of the learners’ achievement level in the education system. In addition, it is used to make performance comparisons among the regions of Namibia, to further identify schools/regions in need of interventions. Namibia has 14 regions, namely, Erongo, Hardap, Karas, Kavango East, Kavango West, Khomas, Kunene, Ohangwena, Omaheke, Omaheke, Oshana, Oshikoto, Otjozondjupa and Zambezi. Although this study focused on the grade 10 learners’ performance in the Khomas region, the problem of poor academic performance is a national debacle. The poor performance of grade 10 learners has been a big concern over the last few years and in the effort to understand this phenomenon there has been efforts to present models that explain it. Modelling semi-continuous data with the presence of excess zeros has become a common phenomenon in real life situations. Common models such as linear models cannot handle zero-inflated data (Zeileis, Kleiber, & Jackman, 2008).

It was against this background information that this study opted to use Generalized Linear Models (GLM) to analyse the factors that influenced the pass rate of Grade 10 learners in Khomas Region in 2016 final Examination. This models included well known Generalised Linear Models such as Poisson regression, Negative Binominal, Hurdle models and Zero inflated models.

1.2 Problem Statement

The academic performance among the grade 10 learners has become a huge matter of concern in the Namibian schools since independence. The annual release of the JSC
Examination results of 2006, 2007 and 2008 revealed that the general poor performance of the schools in most regions, including the Khomas region (Ministry of Education, 2010). For the Khomas region, a fluctuation trend was observed in the overall grade 10 performances since 2012. In 2012, the region’s overall performance was ranked 6th out of the 14 regions of Namibia, while in 2013 it was ranked 8th, and went further down as far as 12th in 2014. However, in 2015, the region’s overall performance went up slightly to 10th position but has never retained its 2012 position. Last year (2016), it remained unchanged at the 10th position (DNEA, 2016). Interestingly, the Khomas region, despite being the location of the capital with numerous private schools, the overall grade 10 performance is appalling. Learners in the Khomas schools seem to have additional and improved facilities as well as equipment needed to learn and perform better than their counterparts in other regions. However, despite all this, learners are performing poorly.

For these reasons, this study sought to find out the factors affecting the pass rate in this region. In addition, the relationship between the performance rate and the types of school (private or government) was examined. Furthermore, possible strategies that could be implemented to improve the grade 10 learners’ performance in the Khomas region were identified.

1.3 Objectives of the study

The main goal of the study was to investigate factors affecting the pass rate of grade 10 learners across schools in the Khomas region, using the Junior Secondary Certificate (JSC) examination results for the year 2016 obtained from the DNEA. The
investigation on factors affecting the pass rate was achieved by addressing the following objectives:

- Exploring the various models that can be potentially applied to analyse the relationships between the pass rate and the demographic and socio-economic variables.
- Applying the best model to analyse the relationship between the pass rate and demographic and socio-economic variables.
- Estimating the effects of demographic and socio-economic variables on the pass rate based on the best model.
- Suggesting measures and strategies that can be used to improve the pass rate of grade 10 learners in the Khomas region.

1.4 Hypothesis of the study

The study assumed that there is a relationship between the ages of the learners, sex, type of school attended (private or government) and location (urban, low, or density), rural and semi-rural (high density) and performance.

1.5 Significance of the study

The findings of this study helped to determine the factors that contribute to the poor performance of the Grade 10 learners in the Khomas region. In addition, measures and strategies that could enhance the learners’ performance would be suggested to the policymakers and stakeholders for consideration. Furthermore, the research was aimed at benefitting the Ministry of Education Directorate in the Khomas region since they would be aware of the major factors that contribute to the learners’ poor performance in their region. The findings would also help improve teaching and learning. Using the
factors that contribute to Grade 10 learners’ poor performance in examinations, a programme of action could be implemented to ensure better Grade 10 learners’ performance. Moreover, this study might also serve as a source of information for future researchers and other scholars in this field.

1.6 Limitations of the study

The study was limited in the sense that it only considered a few (measurable) variables that influence the pass rate of the learners, namely, age, sex, school’s location, and type of school. Factors such as parental socio-economic status, qualification of teachers, and availability of teaching and learning resources were not considered in the study. This might give the impression that the pass rate is influenced by these factors only. Another limitation of this study was that the dataset that was employed showed the Junior Secondary Certificate (JSC) examination results for 2016 only. This was due to the unavailability and lack of access to the Junior Secondary Certificate (JSC) database for the Examination results for the years before 2016, and the bureaucracy involved in getting data from DNEA. However, the findings of this research can be used as basis for further educational research.

1.7 Delimitations of the study

The study was carried out in the Khomas region, focusing only on the grade 10 learners of 2016.
1.8 Organisation of the study

The thesis consists of the following five chapters. Chapter 1 introduces the study by providing the orientation of the study, statement of the problem, objectives of the study, the significance of the study, limitations and delimitations of the study. Chapter 2 discusses the theoretical framework of the study and provides a review of literature. Chapter 3 outlines the methodology which includes the research design, population, sample and sampling procedures, and data collection procedures as well as data analysis. In Chapter 4, the researcher presents the results obtained from the data analysis. Chapter 5 covers the discussions, conclusion and recommendations.
CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

This chapter will present a review of existing literature on academic performance in secondary schools particularly in developing countries. The literature review linked the purpose of the study to various theoretical perspectives and the current nature of variables related to the phenomenon. The literature review also focused on different approaches of models to measure academic performance.

2.2 Theoretical approaches of academic performance

Constructivism: In contrast to didactic approaches such as behaviourism and programmed instruction, constructivism states that learning is an active, contextualized process of constructing knowledge rather than acquiring it (Olusegun, 2015). Knowledge is constructed based on personal experiences and hypotheses of the environment. Learners continuously test these hypotheses through social negotiation. Each person has a different interpretation and construction of knowledge process. The learner is not a blank slate (tabula rasa) but brings past experiences and cultural factors to a situation. Olusegun further elaborated that learners learn by fitting new ideas together with what they already know, hence he believes that learning is affected by the context in which it is taught. He further concluded that learning is active rather than inactive and for constructivism to be effective the learners should be exposed to an environment where they can acquire as much information as possible.

Behaviourism: Is regarded as the direct opposite of constructivism (Olusegun, 2015) and is based on the work of Skinner (1953) who believed that the information itself is
knowable outside the bound of any human mind and any individual interpretation of knowledge can be said to be either correct or incorrect. While much of the early work in formal instructional design derived from behaviourism (objectivism), modern academics have come to accept that learning environment which more closely match the needs of constructivist learning may be more active (Olusegun, 2015).

2.3 Factors leading to poor performance in secondary schools

Poor academic performance is most commonly determined by combining demographic, socioeconomic and environmental factors such as the parents’ educational level, occupational status and income level. It is believed that a low socioeconomic status negatively affects the academic achievement of learners in secondary schools (David, 2014).

David (2014) further elaborated that learner performance is dependent on the socioeconomic background (SEB), sex, school location, school type and learner type. His study based is facts on the findings in Sumbawanga District in Western Tanzania. Considering the fact that the physical geographical location of most secondary schools in the Sumbawanga District is rural, and the physical infrastructure is poor and limited, the communities might be affected by low socio-economic which influence academic performance (David, 2014).

Several other studies by Obiero, Mwebi, & Nyang'ara, (2017) and Owoeye & Yara (2011) have been carried out to identify and analyse the numerous factors that affect academic performance in various centres of learning. Their findings identified factors such as the learners’ efforts, literacy level of parents’ education, parental involvement
(Jeyness, 2012); self-motivation, the age of learners, learning preferences (Obiero, Mwebi, & Nyang’ara, 2017); class attendance and entry qualifications as factors that have a significant effect on the learners’ academic performance in various settings. West and Jones (2013) define education as the process of teaching or training and learning in a school or college to improve knowledge and development skills. Pursuing this line of thought, Obiero, Mwebi and Nyangara state that education for sustainable development has come to be seen as a process of learning how to make decisions that consider the long-term future of the economy, ecology and equity of all communities. Hence, it was critical for this study to examine some of the factors that influence academic performance. The choice of the factors reviewed was based on their importance to the current study.

According to Jeyness (2012) poor academic performance is when the examinee/testee produces work that is below the expected standard. Similarly, Asikhia, (2010) describes poor academic performance as any performance that falls below a desired standard. Okoye (1982), further defines the poor academic performance of a candidate in a learning situation as one in which the candidate fails to attain a set standard of performance in a given evaluation exercise such as a test, an examination or series of continuous assessments as such, thus it is vital for this study to find out why learners fail to attain the required results.

The influence of age and sex on academic performance has been investigated in a number of studies with widely differing conclusions. Research has also shown that men perform better than women in certain settings while women outperform men in other settings (Sommerville & Singaram, 2018). Also to be considered is the
learners’ age. Scholarly observations show that recent changes in educational policies around the world have led to an increase in the number of mature-age admissions in educational institutions (Sommerville & Singaram, 2018).

The relationship between sex and the academic achievement of learners has been contested. However, a gap between the achievement of boys and girls has been found, with girls showing better performance than boys in certain instances (David, 2014). According to Considine and Zappala (2012), the educational performance in school is also influenced by the learner’s sex. Reviews of the evidence by Considine and Zappala (2012), suggest that boys suffer an educational disadvantage relative to girls, especially in terms of performance in literacy. The two scholars further indicate several explanations for this increasing sex gap which include: biological differences; sex biases (such as reading the fact is seen as not being masculine); teaching, curricula and assessment (for instance less structured approaches to teaching grammar) may have weakened boys (Considine & Zappala, 2012).

According to Jeyness (2012) poor academic performance is most commonly determined by combining demographic, socioeconomic and environmental factors such as the parents’ educational level, occupational status and income level. In addition, the low Socio-Economic Status (SES) negatively affects academic achievement of learners in secondary schools (Hansen & Mastekassa, 2013). While a positive relationship between self-motivation and academic performance has been established, the effect of the family’s income and parents’ level of education on academic performance was far from being unraveled without equivocation. The
Socioeconomic status of parents showed moderate to strong relationships with academic performance.

A study conducted by Orlu (2013), among six hundred teachers and learners aimed at establishing the environmental influence on the academic performance of secondary school learners, it was found that the school environment has a significant influence on academic performance, and that its location can affect learners’ performance. For example, when a school is situated in a noisy area like an airport or in the heart of a city where activities disrupt the teaching and learning of the learner, one would not expect the learners to do well academically. In fact, noise in any learning environment interferes with the teaching and learning process.

Overcrowding is another factor that affects the teaching and learning environment. Chuma (2012) observes that overcrowding in classrooms makes it difficult for pupils to write. The teacher is also unable to move around the class freely to assist needy pupils and this affects the teaching-learning process. This means that crowded classroom conditions not only make it difficult for learners to concentrate but inevitably limit the amount of time teachers can spend on innovative teaching methods such as cooperative learning and group work (Chuma, 2012).

This is a result of the parents’ levels of investment in their children’s education which determine their level of purchasing capacity. Parents’ occupations also influence the learners’ achievement in academic work. Learners’ academic achievement is negatively correlated with the low level of the parent’s income which
hinders the individual from gaining access to sources and resources of learning (Jeyness, 2012).

A comparison study between government schools and private schools by Goval (2006) in Orissa city in India, highlighted differences between private and government school performances and found that poor performance in government schools is attributed to poor school resources, poverty and the illiteracy of parents. He further notes that teachers’ absence and inactivity in government schools contributed immensely to poor performance and concluded that the private schools’ performed better due to the following factors: they have higher teacher attendance and activity, the teachers get a better salary compared to government school teachers, and they have smaller class sizes (Goval, 2006).

Given the higher probability that private schools require parents to pay fees, the social background of learners in private and public schools will vary, especially in terms of the occupational, educational and financial status of both parents. Consequently, more learners from a more favourable background tend to go to private schools, which in turn might improve the social composition of the school population. More learners of a favorable background will have more opportunities of reaching higher levels of scholastic achievement, both as a result of a higher level at the start of secondary school, and of better teaching and learning conditions (especially more teaching, due to a lower level of non-academic disturbances). This will promote a potentially better reputation of academic quality for private schools in comparison to public schools, thus attracting different learners (Goval, 2006).
In a study carried out in the United States (U.S) by Lubienski & Lubienski (2006), they compared academic achievement between private and public schools. One of the major findings from this study showed that private schools scored higher grades than public schools. In another related study conducted by Braun, Jenkins and Grigg (2006), they compared the performance of learners in private and public schools in both reading and mathematics for grades 4 and 8. The latter’s results showed that the private schools performed better than the public schools not only in reading and mathematics, but also in the two grades (4th & 8th) involved in the study. The overall findings indicated that the average private school mean score was higher than the average public schools mean score, and that the difference was statistically significant.

Further research by Considine & Zappala (2012) also revealed the importance of the type of school a child attends in influencing educational outcomes. They stated that learners attending private non-Catholic schools were significantly more likely to stay in school than those attending state schools.

Another study carried in Sydney, in Australia by Buckingham (2014) also revealed that the learners from private schools are also more likely to achieve higher end-of-school scores. While school-related factors are important, there is again a direct link to SES, as private schools are more likely to have a greater number of learners from high SES families and they select learners with stronger academic abilities and have greater financial resources (Buckingham, 2000). This situation shows that the poor performance was caused by various circumstances including the unconducive environment for teaching and learning, low teacher’s morale and low motivation.
despite the increase in the number of qualified teachers as the SEDP indicates. Such evidence shows that there is a disparity in examination performance in terms of the school types (Buckingham, 2000).

Based on the controversial findings on the influence of school type, sex and location on learners’ academic performance in the literature, the present study will ascertain whether school type, sex, age differences and location will significantly influence the secondary school learners’ academic performance.

2.4 Performance of learners in secondary school

David (2014) undertook a study to assess factors that influence the academic performance of learners of selected secondary schools in Sumbawanga District, Tanzania. He used a Binary logistic regression model for inferential analysis to determine the impact of some learners’ home based (parents) and environmental factors on the learners’ failing (Grade D and F) or passing (Grade A, B and C) in continuous assessment examinations. The variables David (2014) used were: parent’s income, parents’ educational level, occupation of the parent, provision of meals in school, truancy, peer group, availability of laboratory equipment, walking distance to school, lack of English language competence, inadequate teaching and learning materials, inadequate number of teachers and unavailability of library facilities.

The model fitted in David (2014) was:

\[
\text{Logit}(y = 1) = \log \left( \frac{p}{1-p} \right) = \alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \\
\beta_6X_6 + \beta_7X_7 \ldots + e_i
\]  

(2.1)
\[ y = 1 \text{ if a learner passes the examinations and } 0 \text{ if a learner fails the examinations.} \]

\[ \alpha = \text{Constant term showing the interception of factors influencing learners’ poor performance.} \]

\[ \beta_i X_i = \text{parameter for independent variables namely, the parents’ educational level, occupation, income, provision of meals in school, truancy, peer group, competence in English language of instruction, walking distance, teaching and learning materials, inadequate laboratories, libraries and availability of teachers.} \]

Thus, David (2014)’s study concluded that some variables examined were statistically significant in influencing the poor academic performance of secondary school learners. These include the income of parents at; \((p < 0.05)\), walking distance \(p < 0.05\) and availability of laboratories \(p < 0.01\). In view of the home based factors, the low income of the parents and long walking distance to school were observed to be the cause of an increase in the poor academic performance of a secondary school learner. The lack of laboratories in secondary schools formed the school environmental factor which significantly reduced the learners’ academic performance (Manganga, 2016).

Parents are responsible for taking care of their children’s education expenses. They are enforced to have adequate sources and resources of funds to sponsor their children not only to cover education expenses but also provide basic needs to their families. Low incomes of parents may lead to failure to pay school fees, buy school uniforms and other scholastic requirements needed for learners, hence resulting in learner truancy; a factor that also links to poor academic performance (David, 2014). However the study carried out by Suleman and Hussain (2014) found English language competence, parents’ education level, occupation of the parent, provision of meals in school,
truancy, peer group, teaching and learning materials, libraries and availability of teachers as not statistically significant.

Owoeye & Yara (2011) conducted a study examining how the location of schools relates to the academic performance of learners in Ekiti state of Nigeria between 1990 and 1997. They analysed their data using mean and t – test. The results showed that there were significant differences between learners’ academic achievement of rural and urban secondary schools in senior school certificate examinations \( t = 2.73, p < 0.05 \). The study proved that learners in urban areas had a better academic achievement than their rural counterparts (Owoeye & Yara, 2011).

Amukowa and Atancha (2013), also conducted a study analysing factors that lead to poor performance in Kenya’s certificate of secondary examination in Embu District in Kenya. They used the Linear multiple correlation and multiple regression to analyse the data on school inputs.

a) Linear multiple correlation established the extent to which the independent variable inputs affect the dependent variable (performance). The correlation ranges from – 1.0. to 1.0., which means a correlation can be positive or negative.

b) Multiple Regression Analysis was also used to show the individual effect of each independent variable on the dependent variable. The variables used were the teacher – pupil ration \( X_1 \), learner average admission score \( X_2 \), head teacher qualification and experience \( X_3 \), laboratory expenditure \( X_4 \), instructional material supplied\( X_5 \), teacher’s qualification and experience \( X_6 \).
The equation for the regression analysis was:

\[ P = a + BX_1 + CX_2 + DX_3 + EX_4 + FX_5 + GX_7 \ldots \]  

(2.2)

Where \( P \) = Performance  
\( a \) = Constant and \( B, C, D, E, F, G \) = were the regression coefficients.

Results from the Linear multiple correlations showed that the instructional materials supplied \((X_5)\) and teacher qualification and experience had the highest correlation coefficient with performance \((X)\). Both correlations were significant at 0.001 significance level in a one tailed test. Furthermore, the results revealed that schools which had supplied more instructional materials and had qualified and experienced teachers performed better in the national examinations. There was a negative correlation between the teacher – pupil ratio \((X_1)\) and performance. The correlation coefficient was \(-0.078\). The result from Multiple Regression (stepwise regression analysis) showed that the teacher’s experience was not statistically significant in the regression model at 0.05 significance level in a two tailed test (Amukowa & Atancha, 2013).

Another study assessing the factors influencing the learners’ examination performance in Maswa District, was conducted by Lugayila (2014). The Analysis of Variance (ANOVA) technique was used to compare the means of the variables and the Pearson correlation coefficient under study. Variables used included school ownership (public or private), school type, learning environment, learners readiness, distance to and from the school, teachers motivation, teaching and learning methods, teacher’s commitment monitoring and evaluation, learners book ratio, learners’ desk ratio, learners - teacher ratio, learners - pit latrine ratio, learner – classroom ratio, learners’ attendance,
teachers’ use of allocated time for teaching and the accessibility of library and laboratory services.

The results from ANOVA revealed an insignificant difference between the school ownership and examination performance at \( (p > 0.05) \). This means that between the two school ownership categories: public schools and private schools, the performance did not differ significantly.

The results at the learners’ school type in Analysis of Variance (ANOVA) showed that school types had no significant difference as at level 0.533 or \( (P > 0.05) \). It was concluded that there was no significant difference between the school types on average examination performance as at \( (P > 0.05) \).

The Pearson correlation coefficient was used to determine the relationship existing between the learners’ examination performance and the distance to and from the school. The results showed that performance and the distance to/from the school are negatively correlated (-0.789), and their negative association was also significant as at \( (P < 0.05) \). Thus, the current examination performance in Maswa District has been negatively influenced by the distance to/from the school. The findings revealed that the more the distance to/from the school the poorer the examination performance.

The analysis of the Maswa District further showed that some variables such as the learning environment, learners’ readiness, distance to and from the school, teachers’ motivation, teaching and learning methods, teachers’ commitment monitoring and evaluation, learners’ book ratio, learners’ desk ratio, learner - teacher ratio, learner -
pit latrine ratio, learners’ – classroom ratio, learner’s attendance, teachers’ use of allocated time for teaching accessibility to the library and laboratory services were observed to be insignificant in examination performance as at \( P > 0.05 \) level. This means that there was no relationship between the variables identified with the learners’ examination performance.

Yusuf & Adigun (2010) examined the influence of school type, gender and location on the learners’ academic performance in Ekiti state secondary schools. They used a t-test statistics to determine the significance of variables such as school type, sex and school location on academic performance. The study hypothesised that school type would not significantly influence the learners’ academic performance. The results showed that the t-calculated value of 1.08 was lower than the t-critical value of 1.7. Thus, null hypothesis was not rejected. This essentially means that school type did not significantly influence the learners’ academic performance. For school sex, the t-calculated value of 0.37 was lower than the t-critical value of 1.7. Hence, the null hypothesis was not rejected. The conclusion drawn was that sex did not significantly influence the learners’ academic performance. The same trend was observed for school location where the t-calculated value of 0.68 was lower than the t-critical value of 1.7. Hence, the null hypothesis was not rejected. Once more, school location did not significantly influence the learners’ academic performance (Yusuf & Adigun, 2010).

Considine & Zappala (2012) also conducted a study on factors influencing the educational performance of learners from disadvantaged backgrounds, which is one of the factors of the current study. They fitted a Binomial Logistic regression to estimate the extent to which individuals, family, behavioural and socio-economic factors
influence the learners’ achievement. The results from the Wald test of significance showed that the coefficients were statistically significant for sex, ethnicity, and parental education. Family structure (i.e. two-parent vs. one-parent family), the main source of family income and geographical location did not significantly predict school performance outcomes.

In their study, the two researchers also looked at the odds ratio to see the extent to which defining characteristics affected the likelihood of attaining outstanding results. Their results against sex revealed that girls were 1.7 times more likely to achieve outstanding results compared to boys. The odds of achieving outstanding results also significantly increased for learners with parents who had high education levels. Learners from metropolitan areas were 1.3 times more likely to achieve outstanding results compared to those living in non-metropolitan areas.

2.5 Gaps identified

There are few studies carried out in these areas and those carried out are based on factors such as parental level of education, parent’s income, libraries, availability of textbooks and other related education development. Those studies carried out did not identify the determinants of poor academic performance of the learners, such as, age, sex, location of the school and types of school. Most of the studies done focused on factors influencing the performance of learners in secondary school not specifically in grade 10. Nevertheless, the literature review was used to provide general information on factors influencing learners’ performance in general.
2.6 Statistical models for count data

The models that are developed to handle count data are normally the Poisson regression and the negative binomial. However due to excess zeros, the hurdle and zero inflated models become very important models in studies on count data. General linear models although very useful, have limitations, such as when the response is restricted to binary and count and when the variance of the response depends on the mean. However, the Generalized linear models (GLM’s) extend the general linear framework to address both of the above issues (Zeileis, Kleiber, & Jackman, 2008). GLM involves probability that can be expressed in exponential form. Such distributions are members of the exponential family of distributions written as:

\[
    f(y; \theta, \phi) = \exp \left[ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right]
\]

(2.3)

• \(a(\cdot), b(\cdot)\) and \(c(\cdot)\) are some functions

• The parameter \(\theta\) is a function of the location parameter of the distribution e.g. the mean. This exponential family of distributions include well-known distributions such as the normal distribution, the Poisson distribution and binomial distribution (Zeileis, Kleiber, & Jackman, 2008).

2.6.1 The hurdle model

The number of subjects passed at grade 10 is a count event. Apart from over dispersion many real life count data exhibit more zero observations than would be allowed for the Poisson model (Zeileis, Kleiber, & Jackman, 2008). The hurdle model is a two-component model that consists of the truncated count component such as a Poisson,
Geometric or Negative binomial which is employed for positive counts and a hurdle component model that model zero against the larger counts (Zeileis, Kleiber and Jackman, 2008). The hurdle model is given by:

\[ Pr(Y = 0) = 1 - \pi, 0 \leq \pi \leq 1 \] (2.4)

\[ Pr(Y = y) = \frac{\pi p(y; \theta)}{1 - p(0; \theta)}. y = 1, 2, ... \] (2.5)

Where \( \pi = Pr(Y > 0) \) is the probability of a nonzero response, \( P(y; \theta) \) is an untruncated, or base, probability distribution with parameter vector \( \theta \), and \( P(0; \theta) \) is the base distribution evaluated at 0, and random variable \( y \) is the count response. There are several approaches to parameter estimation in zero-modified count models, including maximum likelihood, penalized quasi-likelihood, and Bayesian methods. For the uncorrelated hurdle models, parameter estimation proceeds by fitting the two model components separately (Neelon, A. O’Malley, & Normand, 2013). The model selection choice can be based on information-theoretic selection measures, such as the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC), also known as the Schwarz criterion (Neelon, O’Malley and Normand, 2013).

\[ AIC = -2 \ln (L) + 2k \] (2.6)

\[ BIC = -2 \ln (L) + 2k \ln (n) \] (2.7)

where \( k = \) parameter, \( L = \) likelihood and \( n = \) sample size. The model yielding the smallest AIC or BIC is the best model.

We consider models for count responses with excess zeros relative to what standard distributional assumptions, such as the Poisson, can predict. In the literature, ‘zero inflated count data’ refers to data for which a generalized linear model has lack of fit
due to disproportionately many zeros. Such data are common in many applications, especially when many subjects have zero observations. Yet, many also have much larger observations so that the overall mean need not be near zero. An example of a variable that one might expect to be zero inflated is the number of times a subject used medical services in the previous year: some subjects may have a zero observation because of chance, whereas others may have a zero observation because of a ‘doctor avoidance’ phobia.

There is considerable literature on modelling cross-sectional zero-inflated count data, using the hurdle model and the zero-inflated count model (Zeileis, Kleiber, & Jackman, 2008). The hurdle model is a two-part model for count data. One part is a binary model if the response outcome is zero or positive. If the outcome is positive, the ‘hurdle is crossed’. Conditional on a positive outcome, the second part uses a truncated model that modifies an ordinary distribution. For instance, this might be a truncated Poisson distribution or a truncated negative binomial distribution. The hurdle model can handle both zero inflation and zero deflation. A separate strand of the literature pertains solely to zero inflation. With this approach, two types of zeros can occur: one comes from the zero state and the other from the ordinary count distribution state. That is, the relevant distribution is a mixture of an ordinary count model, such as the Poisson or negative binomial, with one that is degenerate at zero (Zeileis, Kleiber, & Jackman, 2008). Such zero-inflated count models are more natural than a hurdle model when it is reasonable to think of the population as a mixture, with one set of subjects that will have only a zero response and other subjects that may have a zero response. A good example is the use of medical services example mentioned earlier.
Compared with the substantial literature on cross-sectional zero-inflated count data, few papers have discussed the modelling of clustered, correlated observations, such as occurs with longitudinal data. Dobbie and Welsh (2001) applied marginal models using the generalized estimating equations approach for both parts of a hurdle model. If a within-subject effect is the focus of the study, a random effect approach is natural. Hall (2000) extended the Lambert (1992) zero-inflated Poisson (ZIP) model to handle longitudinal data, adding a random effect to account for the within-subject dependence in the Poisson state. However, Hall’s model does not have a random effect for the part of the model determining the zero inflation. In contrast, Yau (2001) proposed adding a pair of uncorrelated normal random effects for the two components of the hurdle model. They used a penalized quasi-likelihood (PQL) approach for model fitting. When the response is observed at several occasions, a high positive outcome at one time may increase the probability of a positive outcome at another time. These two processes are likely to be correlated and may be influenced by covariates in similar or in different ways. It makes sense to allow correlated random effects in a model, which then requires a more complex fitting process. In addition, the nonzero response may be over dispersed with respect to a truncated Poisson distribution and a truncated negative binomial distribution may be more appropriate. In this study, the researcher developed correlated random effects models. In model fitting, Lin & Breslow (1995) showed that PQL estimators can be biased and inconsistent for highly non-normal binary responses when the random effects have large variance. Rather than PQL, the researcher uses parametric and Non-Parametric Maximum Likelihood (NPML) for model fitting. For a Poisson hurdle model, when the two parts of the hurdle model have the same covariates, Lin & Breslow (1995) further considered a special type of model
that can be used to test for zero inflation. In addition, they considered a simpler approach using a single model – a cumulative logit model with random effects.

Hurdle models can be extended to the regression setting by modelling each component as a function of \( x \) (Mullahy, 1986):

The general form of the hurdle model likelihood function is:

\[
L = \prod_{i \in \Omega_0} (1 - F_1(\beta_1)) \prod_{i \in \Omega_1} \frac{f_2(y, \beta_2)F_1(\beta_1)}{F_2(\beta_2)}
\]  

(2.8)

Where \( \Omega_0 = \{ i | y_i = 0 \} \), and \( \Omega_0 \cup \Omega_1 = \{ 1, 2, \ldots, N \} \).

Taking the natural logarithm of both sides and rearranging terms, the log likelihood can be written as:

\[
\ln(L) = \sum_{i \in \Omega_0} \ln(1 - F_1(\beta_1)) + \sum_{i \in \Omega_1} \ln(F_1(\beta_1)) + \sum_{i \in \Omega_1} \left[ \ln(f_2(y, \beta_2)) - \ln(F_2(\beta_2)) \right]
\]  

(2.9)

2.6.1.1. The Poisson hurdle model

According to Mullah (1986) the probability mass function is:

\[
Pr(Y = y) = \begin{cases} 
\pi, & y = 0 \\
1 - \pi, & y = 1, 2, 3, \ldots
\end{cases}
\]  

(2.10)

The zero-truncated Poisson process has the probability mass function

\[
Pr(Y = y | Y \neq 0) = \begin{cases} 
\frac{\lambda^y}{(e^\lambda - 1)^y}, & y = 1, 2, 3, \ldots \\
0, & otherwise
\end{cases}
\]  

(2.11)
Thus, the unconditional probability mass function for $Y$ is

$$\Pr(Y = y) = \begin{cases} 
\pi, & y = 0 \\
(1 - \pi) \frac{\lambda^y}{(e^\lambda - 1)y!}, & y = 1, 2, 3, \ldots
\end{cases}$$

(2.12)

And the log likelihood for the $t^{th}$ observation, assuming the observations are independently and identically distributed, is

$$\ln L(\pi_t, \lambda_t, y_t) = \begin{cases} 
\ln \pi_t, & y = 0 \\
\ln \left\{ (1 - \pi_t) \frac{\lambda_t^{y_t}}{(e^{\lambda_t} - 1)y_t!} \right\}, & y = 1, 2, 3, \ldots
\end{cases}$$

(2.13)

If we model $\pi_t$ using the complementary log-log link and $\lambda_t$ using the log link, with a little algebra we have:

$$\pi_t = e^{-e^{x_t^1}\beta_1}$$

(2.14)

And

$$\lambda_t = e^{x_t^2\beta_2}$$

(2.15)

The assumptions for Poisson regression are (Mullah, 1986):

- **Y-values are counts**: If your response variables aren’t counts, Poisson regression is not a good method to use.

- **Counts must be positive integers**: i.e. whole numbers 0 or greater (0, 1, 2, 3...k).

  The technique will not work with fractions or negative numbers, because the Poisson distribution is a discrete distribution.

- **Counts must follow a Poisson distribution**: Therefore, the mean and variance should be the same.

- **Explanatory variables must be continuous, dichotomous or ordinal.**

- **Observations must be independent.**
2.6.1.2. Hurdle negative binomial regression

In negative binomial regression, the mean of $y$ is determined by the exposure time $t$ and a set of $k$ regressor variables (the $x$’s). The expression relating to these quantities is:

$$\mu = \exp(\ln (t_i) + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki})$$  \hspace{1cm} (2.16)

often $x_1 = 1$. In which case $\beta_1$ is called the intercept. The regression coefficients $\beta_1, \beta_2, \ldots, \beta_k$ are unknown parameters that are estimated from a set of data. Their estimates are symbolised as $b_1, b_2, \ldots, b_k$.

Using this notation, the fundamental negative binomial regression model for an observation $I$ is written as:

$$Pr (Y = y_i | \mu_i, \alpha) = \frac{\binom{y_i + \alpha - 1}{y_i}}{\Gamma(y_i + \alpha) \Gamma(\alpha - 1)} \left(\frac{\alpha - 1}{\alpha - 1 + \mu_i} \right)^{\alpha - 1} \left(\frac{\mu_i}{\alpha - 1 + \mu_i} \right)^{y_i}$$  \hspace{1cm} (2.17)

2.6.2 Zero-inflated count models

An alternative approach for modelling zero-inflated data is the zero-inflated count model proposed by Lambert (1992). This model assumes that data are from a mixture of a regular count distribution, such as the Poisson distribution, and a degenerate distribution at zero. The EM algorithm or the Newton–Raphson method can be used to obtain the ML estimates. Compared to the hurdle model, this model is more complex to fit, as the model components must be fitted simultaneously.
2.6.2.1 A zero-inflated negative binomial regression model (With hidden Markov chain)

Wang & Alba (2006) consider a random variable $Y$ of event counts with a piece of data set of $k$ subjects:

$\{(y_{ij}, x_{ij}); i = 1, ..., k, j = 1, ..., n_i \}$ where $y_{ij}$ is observed event counts for subject $i$ during the $j$th period, associated with a vector of covariates $x_{ij}$, and the total sample size $n = \sum_{i=1}^{k} n_i$. The proposed model assumes that:

(1) For observed event counts $y_{ij}$ for subject $i$ during period $j$, there corresponds a partially observed binary random variable, $S_{ij}$, representing the state of a two-state discrete time Markov chain with $S_{ij} = 1$ when $y_{ij} > 0$ and $S_{ij} = 0$ or 1 when $y_{ij} = 0$;

(2) The partially observed binary random vector $(S_{i1}, S_{i2}, ..., S_{in})$ for subject $i$ follows the two-state discrete time Markov chain with transition probabilities defined by:

\[
\begin{align*}
\Pr(S_{ij} = 0|S_{i(j-1)} = 0) &= p_{00}, \Pr(S_{ij} = 1|S_{i(j-1)} = 0) = p_{01} \\
&= 1 - p_{00} \\
\Pr(S_{ij} = 1|S_{i(j-1)} = 1) &= p_{11}, \Pr(S_{ij} = 1|S_{i(j-1)} = 1) = p_{10} \\
&= 1 - p_{11}
\end{align*}
\]

(2.18) (2.19)

Where $p_{00}, p_{01}, p_{10}$ and $p_{11}$ are unknown parameters; and

(3) Conditional on $S_{ij} = 1$, observed count $y_{ij}$ follows a negative binomial distribution with

\[
\begin{align*}
f_{1}(y_{ij}|x_{ij}, \alpha, \beta, S_{ij} = 1) = \\
\frac{\Gamma(y_{ij}+\alpha^{-1})}{\Gamma(\alpha^{-1})} \left( \frac{\alpha \lambda_{ij}}{1+\alpha \lambda_{ij}} \right)^{y_{ij}} \left( \frac{1}{1+\alpha \lambda_{ij}} \right)^{\alpha^{-1}}
\end{align*}
\]

(2.20)
Where $\lambda_{ij} = \exp(\beta'x_{ij}, \beta = (\beta_1, ..., \beta_d)$ is an unknown parameter vector and $\alpha \geq 0$ is the dispersion parameter; conditional on $S_{ij} = 0, y_{ij} = 0$, i.e.

$$f_0 = (y_{ij}|S_{ij} = 0) = \begin{cases} 1, & \text{if } y_{ij} = 0 \\ 0, & \text{if } y_{ij} > 0 \end{cases}$$

(2.21)

### 2.6.3 Hurdle model versus ZIP model

The ZIP model is suitable only for handling zero inflation. However, the hurdle model is also suitable for modelling zero deflation. In fact, when a data set is zero deflated at a level of a factor, the estimate of the corresponding parameter in the first part of the ZIP model is 1, so that the fit has no zero inflation at that level. The hurdle model does not have this problem.

Min and Agresti (2005) used two simple simulations to study this potential problem with the ZIP model. The first simulation assumed a hurdle model with Poisson $g_2$, at which there was zero deflation at one setting of the predictor but the entire data set, ignoring the covariate tended to be zero inflated. Not surprisingly, the estimate of the predictor for fitting the ZIP model was highly unstable (Min and Agresti, 2005). However, when both van den Broek (1995) and Jansakul & Hinde (2002) used score tests on our 1000 simulated data sets, both tests revealed evidence of zero inflation ($P-value < 0.05 \text{ in each case}$). This simulation study shows that even when a test shows significant evidence of zero inflation, the ZIP model may still not be suitable to fit the data. More relevantly, the second simulation assumed a ZIP model, for which the hurdle model is also valid.

Another case simulated by Lambert (1992) where 1000 data sets were generated, in which each data set had $n = 200$ observations with a covariate $x_i$ which is binary,
taking value 0 for 100 cases and 1 for the other 100 cases. With these choices, on average, 51% of the responses were zeros, and 22% of those zeros were generated by the Poisson distribution. For this assumed ZIP model, the hurdle model with Poisson $g_2$ also holds. The second part of the hurdle model has the same parameters as the second part of the ZIP model (Lambert, 1992).

Moreover, as quoted in Chin and Quddus (2003), the Zero inflated and hurdle models are indistinguishable using goodness of statistics. Both are used if the data contains excess zeros, however the Hurdle models are useful when the data has excess non-structural (sampling) zeros, whereas the Zero inflated models are useful when the excess zeros are structural (fixed).

### 2.7 Application of the hurdle models and Zero inflated models

#### 2.7.1 The hurdle models

A good example of the application of a hurdle model is found in studies that examine the rates of patient hospitalisation. For instance, many patients will have no hospitalisations, resulting in a count of zero. When the number of zeros is greater or less than expected under a standard count model, the data are said to be zero modified relative to the standard model (Neelon, O’Malley and Normand, 2013). A similar phenomenon arises with semi continuous data, which are characterised by a spike at zero followed by a continuous distribution with positive support (Neelon & Normand, 2013). When analysing zero-modified count and semi continuous data, flexible mixture distributions are often needed to accommodate both the excess zeros and the typically skewed distribution of nonzero values (Neelon, O’Malley and Normand, 2013).
Another case, conducting research in health settings, is to have data where most of the observations have a zero reading. For instance, when observing patients who have been tested for a particular disease compared to those who have not and recording those who have been tested as one and those who have not as zero, there are more likely to be more zeros in the data (Neelon, O’Malley and Normand, 2013).

Similarly, in studies examining outpatient clinic visits, patients who report no visits will be assigned a count of zero. “Count-valued outcomes, like those in the previous two examples, are typically modelled using discrete distributions, such as the Poisson or negative binomial distribution. However, there are times when the proportion of zeros is greater or less than what a standard count distribution would predict, and in such cases, the data are said to be zero modified relative to an ordinary count model”.

“A related phenomenon occurs with semi continuous outcomes, such as medical expenditures, which are characterised by a point mass at zero followed by a continuous distribution for the positive values. When analysing zero-modified count and semi continuous data, parametric mixture distributions known as two-part models are often needed to accommodate both the abundance of zeros and the typically highly skewed distribution of nonzero values” (Neelon & Normand, 2013).

Various two-part models have been developed in recent years to address zero-modified count and semi continuous data, including hurdle models, zero-inflated models, and two-part semi continuous models. While these models vary in terms of their distributional assumptions and parametric forms, they typically incorporate an
underlying two-part structure in which the zero and nonzero observations are modelled through distinct although sometimes overlapping sets of parameters.

### 2.7.2 The Zero inflated models

They Zero inflated models are useful when the data contains excess zeros that are both structural and non-structural (sampling zeros) (Hall, 2000). These models have been extensively used in fields such as econometrics and medical fields (Hall, 2000).

### 2.8 Conceptual framework

This paper hinges on the conceptual framework David (2014) that states that poor academic performance is most commonly determined by combining demographic, socioeconomic and environmental factors such as parents’ educational level, occupational status and income level. Poor pass rates or poor academic achievements can lead to undesirable effects such as disruption of the learning process in the classroom and the whole school as well. As such, learners may fail to progress to tertiary institutions where they can be well equipped for the job market, and this results in high unemployment rates, crime and other socio-economic problems (Younes & Al-Zoubi, 2015).

The article by David (2014) stated that the academic performance of learners in secondary schools over a given period of time may be influenced by socio-economic factors originating from their families, school environment and the learners themselves. In addition, the article notes that socio-economic, socio cultural and socio-political variables influence learners’ performance directly or indirectly, either by the increase or decrease of the number of learners’ average scores of grades A(80-
100%), B (70-79%) and C (50-69) for Pass and grade D (40-49) E (35-39) and F (30-34) for Fail as illustrated in Figure 1. However for Namibia symbols D to G are regarded as pass. These factors have an effect on the learners’ performance in the National Form Four Examinations, in particular, those scoring division I to III grades. By implication, the number of learners qualifying for advanced secondary school and technical college is also affected. The division four and zero scores are also a result of this influence. This conceptualisation shows the complexity of factors affecting learners’ academic performance in the National Form Four Examination results with most variables being interrelated.

![Figure 1: Conceptual Framework](image-url)

Figure 1- Variables Affecting Learner Academic Performance (David, 2014)
CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter provides an outline on how the research process was carried out to get answers to research questions. It includes the research design, population and samples, types of variables and methods of analysis.

3.2 Research design

The study was quantitative in nature. The analysis was based on secondary data obtained from the Directorate of National Examinations and Assessment (DNEA) in the Ministry of Education’s database for the year 2016. The study followed a cross-sectional approach comparing the performance rate of the grade 10 in Khomas region, across several influencing factors.

3.3 Population and samples

The population of this study were the 45 schools in the Khomas region that offer grade 10 education whereby all the 45 schools were used for analysis thus, there was no sampling performed.

3.4 Variables

The dependent for this study was the number of subject passed and the independent variables used were learner’s sex, age, school location (urban, semi-urban, or rural), and type of school (government or private). A learners is expected to pass 6 out of 9 subjects with an average of D symbol (50%).
Table 1: Variables summary

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects passed</td>
<td>The number of subjects were a learner scored points matching symbols A-D.</td>
</tr>
<tr>
<td>Age</td>
<td>Age of the learners (in years) before the next birthday. The age ranges between 15 -29. The oldest learner was found at the private school, Tanben College. Grade 10 learners are expected to be 16-17 years of age.</td>
</tr>
<tr>
<td>Location</td>
<td>The location where the school is located (Rural =0, Low density=1, High density=2).</td>
</tr>
<tr>
<td>Type of school</td>
<td>The type of school were the learner attended and wrote the final exams in 2016(Private =0, State =1)</td>
</tr>
<tr>
<td>Sex</td>
<td>(Female=0, Male =1)</td>
</tr>
</tbody>
</table>

3.5 Methods of analysis

Descriptive statistics was done to graphically explore the data and provide some basic summary statistics. Six models were explored, namely the Poisson regression, Negative binomial, Hurdle Poisson, Hurdle Negative Binomial, Zero Inflated Poisson and Zero- Inflated Negative Binomial models. The hurdle Poisson and the hurdle Negative binomial and Zero-Inflated models are used to account for variables with many zeros, particularly in our case to analyze the variable “number of subjects passed. The Poisson regression is susceptible to over dispersion and the Quasi Poisson as well
as the Negative Binomial are useful when there is over dispersion, which means that the variance is higher than the mean. The analysis of this study did not explore the Quasi Poisson due to some limitations experienced during the R programming. The analysis opted to explore models with the lowest AIC’s. The data was analysed using both the R programming software and the Statistical Packages for Social Sciences (SPSS) software. Several in-built R packages (MASS, pscl and AER packages) were used to handle cases involving the hurdle models and the Zero-Inflated models.

3.6 Data collection procedures

Data for this research was obtained from the DNEA broad sheets provided to each region that contains distribution of subject’s rankings, results per school, school name etc.

3.7 Ethical considerations

The data was collected with the permission of DNEA director. There was a verbal communication with the Khomas Director of Education. However, a formal letter requesting permission for the 2016 JSC examination results for the purpose of this study was written to the Director of Education, Khomas.
CHAPTER 4: RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter presents the results obtained from the study analysis. The chapter is divided into two sections: the Descriptive Statistics section which covers the profile of the dependent and independent variables, and the Inferential Statistics that analyses the relationship between the dependent variable and the independent variables.

4.2 Part I: Descriptive statistics

4.2.1 Profile of the Pass Rate across Sex

Table 2 shows that the failure rate in 2016 from the Khomas Grade 10 results was higher (55%) than the pass rate (45%). Hence, the study’s intention is to explore factors that affect the pass rate in the Khomas region.

Table 2: Pass rate by sex of Grade 10 learners

<table>
<thead>
<tr>
<th>Sex</th>
<th>Count</th>
<th>%</th>
<th>Count</th>
<th>%</th>
<th>Count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail</td>
<td>1527</td>
<td>53.4%</td>
<td>1305</td>
<td>57.0%</td>
<td>2832</td>
<td>55.0%</td>
</tr>
<tr>
<td>Pass</td>
<td>1333</td>
<td>46.6%</td>
<td>985</td>
<td>43.0%</td>
<td>2318</td>
<td>45.0%</td>
</tr>
<tr>
<td>Total</td>
<td>2860</td>
<td>100.0%</td>
<td>2290</td>
<td>100.0%</td>
<td>5150</td>
<td>100.0%</td>
</tr>
</tbody>
</table>
4.2.2 Profile of the number of subjects passed by the grade 10 learners

Figure 2: Bar chart showing the distribution of the number of subjects passed.

The bar chart, Figure 2, reveals that zeros are in excess, compared to the remaining groups. Thus, a hurdle model or zero inflated model was more appropriate to use in the inferential data analysis section to achieve the objectives of this study.
4.2.3 Profile of the types of location

Table 3: Distribution of the types of school across school location

<table>
<thead>
<tr>
<th>Location</th>
<th>Low density</th>
<th>High Density</th>
<th>Rural</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Count</td>
<td>%</td>
<td>Count</td>
<td>%</td>
</tr>
<tr>
<td>Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>51</td>
<td>2.2%</td>
<td>314</td>
<td>11.8%</td>
</tr>
<tr>
<td>State</td>
<td>2296</td>
<td>97.8%</td>
<td>2337</td>
<td>88.2%</td>
</tr>
<tr>
<td>Total</td>
<td>2347</td>
<td>100.0%</td>
<td>2651</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Table 3 shows that of the total number of schools in the high density location, 434 (8.4%) were privately owned while 4716 (91.6%) were state owned. It can also be seen that the schools from the highly populated areas are more (51%) compared to the schools in low populated areas (46%) and Khomas rural (3%). The comparison of rural to low and high density was necessitated by the contrasting nature of Khomas rural schools which are in actual fact farm schools that are far from the city of Windhoek where the other two groups (Low and high density schools are found).
### 4.2.4 Profile of schools and location in Khomas region

Table 4: Profile of schools and location in Khomas Region

<table>
<thead>
<tr>
<th>School</th>
<th>Location</th>
<th>LD</th>
<th>HD</th>
<th>Rural</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CJ BRANDT HS</td>
<td></td>
<td>0</td>
<td>199</td>
<td>0</td>
<td>199</td>
</tr>
<tr>
<td>DAVID BEZUIDENHOUT</td>
<td></td>
<td>0</td>
<td>220</td>
<td>0</td>
<td>220</td>
</tr>
<tr>
<td>EROS GIRLS E</td>
<td></td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>HIGHLANDS CHRISTIA</td>
<td></td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>JAKOB MARENGO T.CO</td>
<td></td>
<td>0</td>
<td>314</td>
<td>0</td>
<td>314</td>
</tr>
<tr>
<td>KHOMAS HS</td>
<td></td>
<td>143</td>
<td>0</td>
<td>0</td>
<td>143</td>
</tr>
<tr>
<td>ST JOSEPH’S RC</td>
<td></td>
<td>0</td>
<td>0</td>
<td>69</td>
<td>69</td>
</tr>
<tr>
<td>ACACIA HS</td>
<td></td>
<td>145</td>
<td>0</td>
<td>0</td>
<td>145</td>
</tr>
<tr>
<td>ACADEMIA SS</td>
<td></td>
<td>154</td>
<td>0</td>
<td>0</td>
<td>154</td>
</tr>
<tr>
<td>A. SHIPENA</td>
<td></td>
<td>0</td>
<td>165</td>
<td>0</td>
<td>165</td>
</tr>
<tr>
<td>AUGUSTINEUM SS</td>
<td></td>
<td>0</td>
<td>181</td>
<td>0</td>
<td>181</td>
</tr>
<tr>
<td>CENTAURUS HS</td>
<td></td>
<td>144</td>
<td>0</td>
<td>0</td>
<td>144</td>
</tr>
<tr>
<td>CONCORDIA COLLEGE</td>
<td></td>
<td>223</td>
<td>0</td>
<td>0</td>
<td>223</td>
</tr>
<tr>
<td>COSMOS HS</td>
<td></td>
<td>181</td>
<td>0</td>
<td>0</td>
<td>181</td>
</tr>
<tr>
<td>DELTA SS</td>
<td></td>
<td>165</td>
<td>0</td>
<td>0</td>
<td>165</td>
</tr>
<tr>
<td>ELDORADO SS</td>
<td></td>
<td>222</td>
<td>0</td>
<td>0</td>
<td>222</td>
</tr>
<tr>
<td>ELLA DU PLESSIS SS</td>
<td></td>
<td>0</td>
<td>230</td>
<td>0</td>
<td>230</td>
</tr>
<tr>
<td>GOREANGAB JSS</td>
<td></td>
<td>0</td>
<td>160</td>
<td>0</td>
<td>160</td>
</tr>
<tr>
<td>GROOT AUB JSS</td>
<td></td>
<td>0</td>
<td>0</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>HAGE GEINGOB SS</td>
<td></td>
<td>0</td>
<td>175</td>
<td>0</td>
<td>175</td>
</tr>
<tr>
<td>School Name</td>
<td>AIC</td>
<td>Log-Likelihood</td>
<td>AIC</td>
<td>Log-Likelihood</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----</td>
<td>----------------</td>
<td>-----</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>HIGHLINE SS</td>
<td>0</td>
<td>131</td>
<td>0</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>HOCHLAND HS</td>
<td>196</td>
<td>0</td>
<td>0</td>
<td>196</td>
<td></td>
</tr>
<tr>
<td>IMMANUEL SHIFIDI S</td>
<td>0</td>
<td>236</td>
<td>0</td>
<td>236</td>
<td></td>
</tr>
<tr>
<td>JAN JONKER AFRIKAN</td>
<td>0</td>
<td>143</td>
<td>0</td>
<td>143</td>
<td></td>
</tr>
<tr>
<td>JAN MOHR PROJECT S</td>
<td>0</td>
<td>203</td>
<td>0</td>
<td>203</td>
<td></td>
</tr>
<tr>
<td>JAN MOHR SS</td>
<td>278</td>
<td>0</td>
<td>0</td>
<td>278</td>
<td></td>
</tr>
<tr>
<td>KHOMAS TURA PROJEC</td>
<td>0</td>
<td>156</td>
<td>0</td>
<td>156</td>
<td></td>
</tr>
<tr>
<td>ROCKY CREST HS</td>
<td>0</td>
<td>138</td>
<td>0</td>
<td>138</td>
<td></td>
</tr>
<tr>
<td>TANBEN COLLEGE</td>
<td>39</td>
<td>0</td>
<td>0</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>WINDHOEK HS</td>
<td>255</td>
<td>0</td>
<td>0</td>
<td>255</td>
<td></td>
</tr>
<tr>
<td>WINDHOEK TECHN. HS</td>
<td>180</td>
<td>0</td>
<td>0</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>2347</td>
<td>2651</td>
<td>152</td>
<td>5150</td>
<td></td>
</tr>
</tbody>
</table>

4.3 Part II: Inferential statistics

Model Comparison

Six different GLMs (Poisson regression, Negative Binomial, Hurdle Poisson, Hurdle Negative Binomial, Zero Inflated Poisson and Zero- Inflated Negative Binomial) were fitted to analyse the effect of the independent variables on the pass rate of grade 10 learners in the Khomas Region. Table 6 shows the obtained AIC and log-likelihood values for these fitted models. The purpose here was to make a comparison that would yield the best model for the analysis of the variability in the pass rate of grade 10 learners in the Khomas Region. The Poisson regression and Negative Binomial models in Table 5 were fitted using the *glm* package in R, while Hurdle Poisson model, Hurdle Negative Binomial, Zero- Inflated Poisson and Zero- Inflated Negative Binomial
models were fitted using the \textit{pscl} and \textit{AER} packages in R. Appendix C shows all the R-codes used in this section.

Table 5: Various \textbf{GLMs} for modelling count data and their AIC and log-likelihood values

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>Log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Poisson regression</td>
<td>26496</td>
<td>-13242.2</td>
</tr>
<tr>
<td>2 Negative Binomial</td>
<td>23771</td>
<td>-11878.35</td>
</tr>
<tr>
<td>3 Hurdle Poisson model</td>
<td>23448.58</td>
<td>$-1.71 \times e^4$</td>
</tr>
<tr>
<td>4 Hurdle Negative Binomial</td>
<td>22995.1</td>
<td>$-1.148 \times e^4$</td>
</tr>
<tr>
<td>5 Zero Inflated Poisson</td>
<td>23448.55</td>
<td>$-1.71 \times e^4$</td>
</tr>
<tr>
<td>6 Zero- Inflated Negative Binomial*</td>
<td>22988.84</td>
<td>$-1.148 \times e^4$</td>
</tr>
</tbody>
</table>

*best model

From Table 5 the lowest AIC value (22988.84) and the highest log likelihood occurred when using the Zero- Inflated Negative Binomial (model 5). The Zero- Inflated Negative Binomial was the best option for explaining the variability in the grade 10 results of the Khomas Region of 2016. The Zero-Inflated Negative Binomial model is discussed below (Sections 4.3.1). Appendix A2 shows all the results of the over dispersion test. Appendix A shows the R-code used to obtain the models’ results. Appendix B shows all the detailed results of Zero-Inflated Negative Binomial model.
4.3.1 Zero-Inflated Negative Binomial model

Table 6: Zero – Inflated Negative Binomial model with logit link

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>P-value</th>
<th>OR(Odd Ratio)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-15.122</td>
<td>0.879</td>
<td>0.000***</td>
<td>0.000</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.015</td>
<td>0.087</td>
<td>0.861</td>
<td>0.985</td>
</tr>
<tr>
<td>Female (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.675</td>
<td>0.039</td>
<td>0.000***</td>
<td>1.964</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High density</td>
<td>0.131</td>
<td>0.255</td>
<td>0.608</td>
<td>1.140</td>
</tr>
<tr>
<td>Low density</td>
<td>-0.788</td>
<td>0.261</td>
<td>0.0033***</td>
<td>0.455</td>
</tr>
<tr>
<td>Rural (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>1.831</td>
<td>0.337</td>
<td>0.000***</td>
<td>6.240</td>
</tr>
<tr>
<td>Private (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***significant at 5 % level of significance.

Table 6 above reveals that the odds of having no subject passed among male learners is 0.985 times lower than the odds of having no subject passed among females. However, the p-value of 0.861 (in Table 6) is larger than 0.05, hence one can conclude that the male children do not necessarily fail more subjects than female learners. A one unit increase in the number of years (being one year older), increases the odds of not passing a subject by 1.964 times. A learner in a low density area has a reduced chance of 0.455 (45.5%) of failing a subject compared to a learner in a rural school. Being in
a state school, increases the chance of not passing a subject by 6.240 compared to a learner in a private school.

Table 7: Zero- Inflated Negative Binomial model with log link

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Standard error</th>
<th>P-value</th>
<th>Expected Rate(ER)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.425</td>
<td>0.219</td>
<td>0.000***</td>
<td>227.011</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>0.022</td>
<td>0.021</td>
<td>0.295</td>
<td>1.022</td>
</tr>
<tr>
<td>Female (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-0.220</td>
<td>0.012</td>
<td>0.000***</td>
<td>0.803</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High density</td>
<td>-0.454</td>
<td>0.065</td>
<td>0.000***</td>
<td>0.635</td>
</tr>
<tr>
<td>Low density</td>
<td>-0.208</td>
<td>0.065</td>
<td>0.001***</td>
<td>0.812</td>
</tr>
<tr>
<td>Rural (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>School type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>0.300</td>
<td>0.051</td>
<td>0.000***</td>
<td>1.350</td>
</tr>
<tr>
<td>Private (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***significant at 5 % level of significance.

Table 7 reveals that a male learner has a higher chance of 1.022 times of passing than a female learner. However, among learners with a positive number of subjects passed, the p-value of 0.295 is larger than 0.05, hence male children do not necessarily pass more subjects than female learners. The value of 0.803 indicates that if a learner’s age increases by one year among, it will lead to the reduction of the number of subjects
passed by a learner by 0.803 times. A learner who is attending school in a highly populated area has a reduced chance of 0.635 of passing subjects compared to a learner in a rural school, given that the learner has a positive number of subjects passed. Moreover, a learner attending school in a low-density area has a low chance of 0.812 times of passing subjects than a learner in a rural area, given that the learner has a positive number of subjects passed. The value of 1.350 indicates that the learner at the state school has a higher chance of 1.350 times of passing more subjects than a learner at the private school, given that the learner has a positive number of subjects passed.

4.4 Discussions

The results revealed that the location of the school, the age of the learner and the type of school she/he attended are the main contributing factors to the number of subjects passed by the learner and ultimately to the overall pass rate of the learner. When compared to standard models such as Poisson and binomials, the proposed generalised linear models performed well in terms of explaining the variability of the number of subjects passed, consequently the overall pass rate. The generalised Zero-Inflated Poisson model provided the most satisfactory fit among the types of fitted models to illustrate the pass rate of grade 10 learners in the Khomas region.

According to (Obiero, Mwebi, & Nyang'ara, 2017), a number of studies have identified and analysed several factors that affect academic performance in various centers of learning. Their findings identify learners’ effort, parents’ education, family income, self-motivation, age of learner, learning preferences (Obiero, Mwebi, & Nyang’ara, 2017) as factors that have a significant effect on the learners’ academic performance in varied settings.
4.4.1 Age

The statistical analysis of the relationship between the learner’s age and learner’s academic performance for this study showed a significant effect on their final grade 10 results. Precisely, this means that the learner’s age had an effect on one’s performance. Specifically, a larger proportion of the older learners’ performance was lower than the younger learners’. This low performance could be explained by the fact that older learners in classrooms tended to have a lower self-esteem in education and hence, the failure to perform well as other learners. Haist et al. (2000) also established that the age and sex of a learner have an influence on academic performance.

4.4.2 Sex

Contrary to most research, this study found that sex was insignificant or had no effect on the overall academic performance of grade 10 learners. Haist et al. (2000) noted that males perform better than females in certain settings while women outperform men in other settings (Haist et al., 2000). This finding may surprise some educators and parents, as male children often seem to be the focus of discussions on sex differences in the classroom (Considine & Zappala, 2012). Boys mature at a slower rate than girls and are therefore less prepared for formalised schooling (Considine & Zappala, 2012). A gap between the achievement of boys and girls has also been found, with girls showing better performance than boys in certain instances (David, 2014). Unlike in the olden days when girls were expected to stay home and do most household chores, nowadays, children do equal work regardless of their sex. In urban areas there is less work to be done by children that is why they give equal attention to their school work.
However, the results from this current study indicate that learner sex was not a factor in learner academic performance for the overall grade 10 results.

4.4.3 School environment

Concurring with Orlu (2013), this study found the school environment to have a significant influence on academic performance. The location of the school affects learners’ performance, for example, when a school is sited in a noisy area like an airport or in the heart of a city where activities disrupt the teaching-learning process. One would not expect learners in such an environment to perform well academically.

Noise, may be defined as anything that interferes with the teaching/learning process (Orlu (2013)).

Chuma (2012) study which observed that overcrowding in classrooms makes it difficult for pupils to write and make it difficult for teachers to apply innovative teaching methods such as cooperative learning and group work concurred with the findings of this study as well.

The study also coincide with Owoeye and Yara (2011) who noted that the small class size promoted individualised attention, which enhanced the teaching-learning process as the teachers were able to identify the weak learners, attend to the individual needs of the pupils and comfortably address the problems the pupils faced in the learning process.
4.4.4 School type

Apart from age, the school type (State/government) was also identified as a significant factor influencing the performance of a learner. Goval (2006), indicates that the sources of the private school’s advantages lie in the following factors:

(a) They have a higher teacher attendance and activity, (b) the teachers get a higher salary than their counterparts in government schools, and (c) they have smaller class sizes. Given the higher probability that private schools request for fees from parents, the social background of learners in private and public schools will vary, especially in terms of the occupational, educational and financial characteristics of both parents. Consequently, more learners from a more favorable background will go to private schools, which in turn might improve the social composition of the school population. More learners of a favourable background will increase the opportunities of reaching higher levels of scholastic achievement, both as a result of a higher level at the start of secondary school, and of better teaching and learning conditions (especially more teaching, due to a lower level of non-academic disturbances). The differences in conditions promote a potentially better reputation of academic quality for private schools in comparison to public schools, thus attracting different learners.

4.5 Conclusion

This chapter presented findings based on all the results obtained from the study’s data analyses. Six GLMs were initially fitted to determine the best estimated model to use for modelling the factors affecting the grade 10 learners’ pass rate. The Zero-Inflated Negative Binomial model was the best model in terms of the AICs and Log likelihood. The Zero-Inflated Poisson, the two hurdle models, the Hurdle Poisson and the Hurdle
Negative Binomial were the next best models. The age and school type factors were significant in all the models used in this study. The sex variable was insignificant in p-and this contradicts the assumption by many parents who believe that boys do better in secondary schools. As expected, the low-density schools performed better than the rural schools. The state schools’ failure rate was higher a than expected, however the count model revealed that there were more students who had passed more subjects in state schools compared to private schools. This difference can be partly explained by the high number of state schools compared to the low number of private schools that enrolled for the National Junior Secondary Examinations in 2016. The high-density schools also outperformed the rural schools, although this variable was found to be insignificant.
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction
This chapter presents the summary, conclusion and recommendations for the topic under study. The main objective of the study was to investigate factors affecting the pass rate of grade 10 learners across schools in the Khomas region, using the Junior Secondary Certificate (JSC) examination results for the year 2016 obtained from the DNEA. The study further explored the various models that can be potentially applied to analyse the relationship between the pass rate and the demographic and socio-economic variables. The Zero-Inflated Negative Binomial was chosen as the best model to analyse the relationship between the pass rate and, demographic and socio-economic variables, as well as the effects of the demographic and socio-economic variables on the pass rate.

5.2 Conclusions
This study concluded that the poor performance of the grade 10 learners is a challenge in the Khomas region, Windhoek. It was found that the age of the learners, location and school had an effect on the performance of learners. The study also established that the sex of the learners had no impact on their overall performance. As such, the study concurred with Considine & Zappala’s (2012) research on factors influencing the educational performance of learners from disadvantaged backgrounds. Although they fitted a Binomial Logistic regression to estimate the extent to which individual, family, behavioural and socio-economic factors contribute to learners’ achievement, the results from the Wald test revealed that the coefficients were statistically significant for sex, ethnicity, and parental education. However, in terms of the location, the study revealed that schools in low-density areas performed better than rural
schools. They concluded that the geographical location did not significantly predict school performance outcomes. This study however yielded the same results in the case of learners’ performance in highly populated areas, where the variable was insignificant in the Zero Inflated Poisson model.

Although their results on sex revealed that girls were 1.7 times more likely to achieve outstanding results compared to boys, Considine & Zappala’s (2012) study revealed that sex was insignificant in the Hurdle models as well as in the Zero Inflated Model. According to Considine & Zappala (2012), the learners from metropolitan areas were 1.3 times more likely to achieve outstanding results compared to those living in non-metropolitan areas. This corresponded to the findings of this study that learners who attended school in the city performed better than the learners from rural schools.

5.3 Recommendations

Based on the findings of this study, the following recommendations are made that could be implemented to alleviate the poor performance of the Grade 10 learners on the national examinations:

a) State owned schools must have the same privileges, infrastructures and teaching methodologies as private schools.

b) In order to improve the density per school, an emphasis should be put on building more schools in the area so that classrooms are not overcrowded.

c) Rural schools should be given the same attention as low density schools. The current bush allowance should be improved to attract qualified teachers to rural schools.
d) The state schools should bring the teacher-learner ratio of 1:40 in secondary schools on par with the private schools’ ratio of 1:25. This enables teachers in state schools to give more attention to the slow learners. It will also ease the marking load which consume teachers’ time of preparation for lessons that will promote effective teaching and learning.

e) The Ministry should streamline the education system to match the best systems in the world. Figure 8 illustrates a good classroom atmosphere, where the class teacher is fully engaged with the learners.

Figure 3: Grade 3 learners of the attached primary school at Okayama city.
References


Appendix A: R-codes

library(MASS)
library(pscl)
library(AER)
data <- read.csv(file.choose(),header=T)
attach(data)
names(data)
y <- cbind(AD)
x1 <- cbind(Points, Male, Age, HD, LD, State)
x <- cbind(Points, Male, Age, HD, LD, State)
poisson <- glm(y~x, family=poisson)
summary(poisson)
dispersiontest(poisson)
dispersiontest(poisson, trafo=2)
negbin <- glm.nb(y~x)
summary(negbin)
Hurdlepoisson <- hurdle(y ~ x | x1, link = "logit", dist ="poisson")
summary(Hurdlepoisson)
Hurdlenegbin <- hurdle(y ~ x | x1, link = "logit", dist ="negbin")
summary(Hurdlenegbin)
geroinflpoisson <- zeroinfl(y ~ x | x1, link = "logit", dist ="poisson")
summary(zeroinflpoisson)
geroinflnegbin <- zeroinfl(y ~ x | x1, link = "logit", dist ="negbin")
summary(zeroinflnegbin)
logLik(poisson)
logLik(negbin)
AIC(Hurdlepoisson )
AIC(Hurdlenegbin )
AIC(zeroinflpoisson )
AIC(zeroinflnegbin )
Appendix B: Test for over-dispersion

> dispersiontest(poisson)

   Overdispersion test

data:  poisson

   z = 18.1, p-value < 2.2e-16
alternative hypothesis: true dispersion is greater than 1
sample estimates:

dispersion
    1.61736

> dispersiontest(poisson, trafo=2)

   Overdispersion test

   data:  poisson

   z = 14.406, p-value < 2.2e-16
alternative hypothesis: true alpha is greater than 0
sample estimates:

   alpha
    0.1145013
Appendix B: Models Results

Model 1: Poisson regression model

Deviance Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-4.0858</td>
<td>-1.748</td>
<td>-0.3793</td>
<td>0.9929</td>
<td>4.5869</td>
</tr>
</tbody>
</table>

Coefficients:

|                | Estimate  | Std.Error | z value | Pr(>|z|)  |
|----------------|-----------|-----------|---------|-----------|
| (Intercept)    | 7.982337  | 0.155385  | 51.371  | <2e-16 ***|
| xMale          | 0.016229  | 0.014864  | 1.092   | 0.275     |
| xAge           | -0.368147 | 0.008471  | -43.459 | <2e-16 ***|
| xHD            | -0.40583  | 0.042859  | -9.469  | <2e-16 ***|
| xLD            | -0.025407 | 0.04294   | -0.592  | 0.554     |

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 17468 on 5146 degrees of freedom
Residual deviance: 13902 on 5141 degrees of freedom

(3 observations deleted due to missingness)
AIC: 26496
Number of Fisher Scoring iterations: 5

Model 2: Negative Binomial model

> negbin <- glm.nb(y~x)
> summary(negbin)
Call:

`glm.nb(formula = y ~ x, init.theta = 1.773578729, link = log)`

Deviance Residuals:

<table>
<thead>
<tr>
<th></th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>-2.4999</td>
<td>-1.1522</td>
<td>-0.2337</td>
<td>0.5377</td>
</tr>
<tr>
<td>Max</td>
<td>3.5555</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Model: Coefficients

|        | Estimate | Std.Error | z value | Pr(>|z|) |
|--------|----------|-----------|---------|---------|
| (Intercept) | 8.01519  | 0.25388   | 31.571  | < 2e-16 *** |
| xMale    | 0.03064  | 0.02654   | 1.154   | 0.248   |
| xAge     | -0.3711  | 0.01342   | -27.663 | < 2e-16 *** |
| xHD      | -0.39148 | 0.0786    | -4.981  | 6.33E-07 *** |
| xLD      | 0.01478  | 0.07941   | 0.186   | 0.852   |
| xState   | 0.02344  | 0.05403   | 0.434   | 0.664   |

(Dispersion parameter for Negative Binomial (1.7736) family taken to be 1)

Null deviance: 7514.5 on 5146 degrees of freedom

Residual deviance: 6234.9 on 5141 degrees of freedom

(3 observations deleted due to missingness)

AIC: 23771

Number of Fisher Scoring iterations: 1

Theta:  1.7736

Std. Err.:  0.0651

2 x log-likelihood:  -23756.6930
Model 3: Hurdle Poisson model

> Hurdlepoisson <- hurdle(y ~ x | x1, link = "logit", dist = "poisson")
> summary(Hurdlepoisson)

Call:

hurdle(formula = y ~ x | x1, dist = "poisson", link = "logit")

Pearson residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-2.3334</td>
<td>-0.8651</td>
<td>-0.2667</td>
<td>0.8596</td>
<td>8.2227</td>
</tr>
</tbody>
</table>

Count model coefficients (truncated Poisson with log link):

|     | Estimate | Std.Error | z value | Pr(>|z|) |
|-----|----------|-----------|---------|----------|
| (Intercept) | 5.362718 | 0.168899 | 31.751  | < 2e-16 *** |
| xMale       | 0.016122 | 0.015273 | 1.056   | 0.29115  |
| xAge        | -0.214942| 0.009247 | -23.245 | < 2e-16 *** |
| xHD         | -0.353386| 0.044268 | -7.983  | 1.43E-15 *** |
| xLD         | -0.124441| 0.044311 | -2.808  | 0.00498** |
| xState      | 0.211402 | 0.035607 | 5.937   | 2.90E-09*** |
Zero hurdle model coefficients (binomial with logit link):

|            | Estimate | Std.Error | z value | Pr(>|z|) |
|------------|----------|-----------|---------|----------|
| (Intercept)| 14.19559 | 0.66031   | 21.498  | < 2e-16 *** |
| x1Male     | 0.01881  | 0.07166   | 0.262   | 0.793    |
| x1Age      | -0.66715 | 0.03211   | -20.777 | < 2e-16 *** |
| x1HD       | -0.44878 | 0.2289    | -1.961  | 0.0499*  |
| x1LD       | 0.46231  | 0.23419   | 1.974   | 0.0484*  |
| x1State    | -0.93151 | 0.13905   | -6.699  | 2.10E-11 *** |

Number of iterations in BFGS optimization: 13

Log-likelihood: -1.171e+04 on 12 Df

**Model 4: Hurdle negative Binomial model**

```r
> Hurdlenegbin <- hurdle(y ~ x | x1, link = "logit", dist ="negbin")
> summary(Hurdlenegbin)
```

Call:

hurdle(formula = y ~ x | x1, dist = "negbin", link = "logit")

Pearson residuals:

<table>
<thead>
<tr>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1.6952</td>
<td>-0.8012</td>
<td>-0.2438</td>
<td>0.6832</td>
<td>8.1079</td>
</tr>
</tbody>
</table>
Count Model coefficients (truncated negbin with log link):

|               | Estimate | Std.Error | z value | Pr(>|z|)   |
|---------------|----------|-----------|---------|------------|
| (Intercept)   | 5.45173  | 0.221     | 24.668  | < 2e-16 ***|
| xMale         | 0.02156  | 0.02099   | 1.027   | 0.30432    |
| xAge          | -0.22149 | 0.01207   | -18.348 | < 2e-16 ***|
| xHD           | -0.4117  | 0.06386   | -6.447  | 1.14E-10***|
| xLD           | -0.16939 | 0.06425   | -2.636  | 0.00838**  |
| xState        | 0.26285  | 0.04848   | 5.422   | 5.88E-08***|
| Log(theta)    | 1.79816  | 0.07058   | 25.477  | < 2e-16 ***|

Zero hurdle model coefficients (binomial with logit link):

|               | Estimate  | Std.Error | z value | Pr(>|z|)   |
|---------------|-----------|-----------|---------|------------|
| (Intercept)   | 14.19559  | 0.66031   | 21.498  | < 2e-16 ***|
| x1Male        | 0.01881   | 0.07166   | 0.262   | 0.793      |
| x1Age         | -0.66715  | 0.03211   | -20.777 | < 2e-16 ***|
| x1HD          | -0.44878  | 0.2289    | -1.961  | 0.0499*    |
| x1LD          | 0.46231   | 0.23419   | 1.974   | 0.0484*    |
| x1State       | -0.93151  | 0.13905   | -6.699  | 2.10E-11***|

Theta: count = 501368.5836

Number of iterations in BFGS optimization: 36

Log-likelihood: -9555 on 15 Df

**Model 5: Zero inflated Poisson model**

```r
zeroinflpoisson <- zeroi nfl(y~x | x1, link ="logit", dist ="poisson")
summary (zeroinflpoisson)
```
Count model coefficients (Poisson with log link):

|         | Estimate | Std.Error | Z-value | Pr(>|z|) |
|---------|----------|-----------|---------|----------|
| (Intercept) | 5.324947 | 0.168033  | 31.69   | < 2e-16 *** |
| xMale    | 0.016307 | 0.015278  | 1.067   | 0.28582  |
| xAge     | -0.212893| 0.009194  | -23.157 | < 2e-16 *** |
| xHD      | -0.363143| 0.044363  | -8.186  | 2.71E-16*** |
| xLD      | -0.132555| 0.044334  | -2.99   | 0.00279 ** |
| xState   | 0.22172  | 0.036516  | 6.072   | 1.26E-09*** |

Zero-inflation model coefficients (Binomial with logit link):

|         | Estimate | Std. Error | r z value | Pr(>|z|) |
|---------|----------|------------|-----------|----------|
| (Intercept) | -14.04243| 0.72195    | -19.451   | < 2e-16 *** |
| x1Male  | -0.01421 | 0.0764     | -0.186    | 0.8524   |
| x1Age   | 0.64372  | 0.03451    | 18.651    | <2e-16 *** |
| x1HD    | 0.3105   | 0.23678    | 1.311     | 0.1897   |
| x1LD    | -0.57605 | 0.24192    | -2.381    | 0.0173 * |
| x1State | 1.2593   | 0.19082    | 6.599     | 4.13e-11 *** |

Number of iterations in BFGS optimization: 18
Log-likelihood: -1.171e+04 on 12 Df

Model 6: Zero- Inflated Negative Binomial model

```
zeroinflnegbin <- zeroinfl(y ~ x | x1, link = "logit", dist = "negbin")
summary(zeroinflnegbin)
```
Count model coefficients (negbin with log link):

|                | Estimates | Std. Error | Z value | Pr (>|z|) |
|----------------|-----------|------------|---------|-----------|
| (Intercept)    | 5.42513   | 0.21922    | 24.747  | < 2e-16 *** |
| xMale          | 0.022     | 0.02101    | 1.047   | 0.29515   |
| xAge           | -0.21989  | 0.012      | -18.326 | < 2e-16 *** |
| xHD            | -0.45372  | 0.06464    | -7.019  | 2.24e-12 *** |
| xLD            | -0.2075   | 0.0648     | -3.202  | 0.00136 ** |
| xState         | 0.30015   | 0.05058    | 5.934   | 2.95e-09 *** |
| Log(theta)     | 1.78478   | 0.07032    | 25.38   | < 2e-16 *** |

Zero-inflation model coefficients (Binomial with logit link):

|                | Estimates  | Std. Error | Z value | Pr (>|z|) |
|----------------|------------|------------|---------|-----------|
| (Intercept)    | -15.12243  | 0.87887    | -17.207 | < 2e-16 *** |
| x1Male         | -0.01523   | 0.0867     | -0.176  | 0.86052   |
| x1Age          | 0.67451    | 0.03904    | 17.279  | < 2e-16 *** |
| x1HD           | 0.13078    | 0.25462    | 0.514   | 0.6075    |
| x1LD           | -0.7875    | 0.26083    | -3.019  | 0.00253 ** |
| x1State        | 1.83078    | 0.33747    | 5.425   | 5.8e-08 *** |

Theta = 5.9583

Number of iterations in BFGS optimization: 42

Log-likelihood: -1.148e+04 on 13 Df
Appendix C: AIC results

> logLik(poisson)
'log Lik.' -11354.09 (df=7)

> logLik(negbin)
'log Lik.' -11878.35 (df=7)

> AIC(Hurdlepoisson )
[1] 23448.58

> AIC(Hurdlenegbin )
[1] 22995.1

> AIC(zeroinflpoisson )
[1] 23448.55

> AIC(zeroinflnegbin )
[1] 22988.84