Application of Linear Mixed Effects Model on Hierarchical Data (KCPE Examination Scores)

Polycarp Omondi Otieno

I56/70521/2013

A research project submitted to the School of Mathematics in partial fullment of the requirements for the Master of Science degree in Social Statistics

July, 2015

Declaration

This research project report is my original work and has not been submitted to any other institution for an academic award.

Signature... Date...............................

Polycarp Omondi Otieno

This research project report has been submitted for examination with my approval as the university supervisor.

Signature... Date...............................

Dr. Nelson Owuor Onyango

Dedication

This project is dedicated to my wife Victoria and son Landon; My parents Charles and Turphenah; Brothers John, Victor and Collins; and Sister Nancy. The moral support and joy I have found in my family during this period is immeasurable.

Acknowledgment

First and foremost my acknowledgment goes to the Almighty God for preserving my life and giving me the strength to go through the two years of my MSc Program.

I would like to pay my most sincere gratitude to my supervisor, Dr. Nelson Owuor Onyango for the insight and guidance he has given to me throughout this work. The mixed effects model being a 'developing'model, comprehending its properties to detail has been very demanding but I thank my him for his patience and understanding on me. The high standards he set motivated me not only to piece up this research work for program completion, but also to ensure that I understood the model I have presented. I humbly remain indebted to him and I believe that there will be opportunity in future to model the complexities involved in higher level hierarchies.

I wish to acknowledge the members of staff of the School of Mathematics under the leadership of Prof. P.G.O Weke for their commitment to the development of the school. To my classmates, I reserve special thanks for the insight they have given me throughout my the masters program.

Finally, I would like to thank my workmates for being very understanding colleagues and sharing the load of my work whenever I needed to have more time on my studies during working days.

Contents

List of Tables

List of Figures

Abstract

Over the last four decades, the mixed effects model has gained prominence in social research in education, health and fields whose data naturally have hierarchical structures. The ability of the mixed effects model to handle unbalanced clustered data as well as providing analysis of within groups and between groups variations have been behind the growing inclination to the use of this model. Prior to its development, the standard linear regression model had been immensely employed in modeling effects or influence of selected factors on the observed phenomena. Hierarchical data structures have subjects nested within groups, a fact that introduces correlation between subjects of the same group and calls for the application of an appropriate model that sufficiently explains the origin of variations.

In this study the linear mixed effects model is applied in the analysis of national examination scores for pupils who sat for that examination in 2013. The number of pupils in each of the sampled schools is variant. The imbalance in the data structure inhibits the application of models like multivariate regression or the ANOVA model and therefore stresses the choice for the linear mixed effects model. The research project builds a suitable linear mixed effects model that explains the distribution of the observed examination scores given pupils'individual characteristics as well as school level characteristics.

The structure of the data used in this study is nested. Insight is provided for the process of choosing fixed and random effects variables and the Akaike Information

Criterion is used in selecting the optimal model used the analysis. The adequacy of the selected model is assessed using the Likelihood Ratio Test which compares the likelihood functions of the selected model and a restricted null model. The Restricted Maximum Likelihood Method has been considered for the estimation of variance components of the model. The Best Linear Unbiased Estimator is employed in the estimation of the fixed effects parameters while the estimates of the random effects parameters are predicted using the Best Linear Unbiased Predictor. Finally, the results of the analysis have been presented with fixed and random coefficients and interpretation given for the results.

Abbreviation

AIC	Akaike Information Criteria
ANOVA	Analysis of Variance
ASAL	Arid and Semi-Arid Lands
BIC	Bayesian Information Criteria
BLUE	Best Linear Unbiased Estimator
BLUP	Best Linear Unbiased Predictor
FPE	Free Primary Education
GER	Gross Enrolment Rate
GLSE	Generalized Least Squares Estimator
HLM	Hierarchical Linear Model
KCPE	Kenya Certificate of Primary Education
KCSE	Kenya Certificate of Secondary Education
KNEC	Kenya National Examination Council
LMER	Linear Mixed Effects Regression
LRT	Likelihood Ratio Test
MDGs	Millennium Development Goals
MLE	Maximum Likelihood Estimate
MOEST	Ministry of Education Science and Technology
OECD	Organization for Economic Cooperation and Development
OLS	Ordinary Least Squares

PIRLS	Progress in International Reading Literacy Study
PISA	Program for International Students Assessment
REML	Restricted Maximum Likelihood Estimate
RESET	Regression Specification Error Test
SACMEQ	South African Consortium for Monitoring Education Quality
SFP	School Feeding Program
TIMSS	Trends in International Mathematics and Science Study
UNESCO	United Nations Education Scientific and Cultural Organization

Chapter 1

Introduction

1.1 The Purpose of Education

According to (King, 1947) "education has a two-fold function to perform in the life of man and in society: the one is utility and the other is culture. Education must enable a man to become more efficient, to achieve with increasing facility the legitimate goals of his life,"and adds, "complete education gives one not only power of concentration, but worthy objectives upon which to concentrate". The thought presented by (King, 1947) is very powerful. Looking at education for what it is and considering the purpose that is argued clearly points to one glaring fact, that without equity, education systems are not efficient and thus do not meet the purpose for which they are established.

(Foshay, 1991) articulates in his persuasion to review school curriculum that "educational purpose have also been widely accepted to develop the intellect, to serve social needs, to contribute to the economy, to create an effective work force, to prepare students for a job or career, to promote a particular social or political system."In a similar thought process to the proposition of King, Foshay acknowledges development of intellect as one of the purposes of education. In my modest consideration, achieving half intellect is very detrimental to a nation since the neglected half may not stand up to serve the social responsibilities that their nations require of them once they graduate from schools. As noted by (OECD, 2013), "increase in the results obtained from PISA increased the likelihood of an adult to contribute meaningfully to the economy of his/her nation. This implies that if nations are not keen to address equity issues in education, then the population that will be available to actively work the economy will be shrunken."

1.2 Introduction

Education and training has been presented by world leaders and scholars as one of the necessary ingredients to economic growth in developed and developing countries. This is a persuasion that is supported by development of education in developed countries especially those that are members of the Organisation for Economic Co operation and Development (OECD). In OECD countries, education is not viewed as an end in itself but as a means to economic prowess. It is for this reason that education assessments have been conducted in such countries and the impact of any improvement made in the levels of achievements associated with economics growth.

In most societies, education is viewed as an avenue to improve not only the economic strongholds of nations but also the social fundamentals. This is seen through the quality of social choices made by the educated citizens across the globe, there is increased range of social choices that are available for the educated compared to the constituency of persons that are not educated. It is suffice to conclude that there is significant relationship between education and economic growth whether one is looking at developed or developing countries.

Results from the OECD's recent Survey of Adult Skills show that highly skilled adults are twice as likely to be employed and almost three times more likely to earn an above-median salary than poorly skilled adults. In other words, poor skills severely limit people's access to better-paying and more rewarding jobs. Highly skilled people are also more likely to volunteer, see themselves as actors rather than as objects of political processes, and are more likely to trust others. Fairness, integrity and inclusiveness in public policy thus all hinge on the skills of citizens. (OECD, 2013)

As evidenced by data from developed and developing countries, access to primary education has grown over the years. The growing numbers of schools both in public and private sectors around the globe is therefore not an accident but a deliberate effort by individual countries to develop their education systems for sustainable development. Majority of citizens around the world desire to acquire basic literacy as they appreciate that education improves their livelihoods and in the same stride they have a chance to build their economies. Apart from increasing the numbers of schools and learning institutions, many countries have reformed the way they look at education. The South African Consortium for Monitoring Education Quality (SACMEQ) for instance carries periodic education assessments in member countries to establish the quality of education programs in each country. Countries participating in the Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS) have increasingly been interested in monitoring achievements of their systems which of course are used in developing the systems in the directions they desire based on lessons learnt from countries whose education systems perform comparatively better.

Through the last decade, countries all over the world rallied to achieve universal primary education through the United Nations led banner of Millennium Development Goals (MDGs). Most countries have since achieved the goal while some are yet to get there, especially in the African continent. While schools opened their gates for school going age children to join the learning process, a new challenge was born. Quality of education, in the eyes of education stakeholders, was put to test in most of the African states, with much of the impacts felt claimed to be in public schools. Quality education is, "Processes through which trained teachers use child-centred teaching approaches in well-managed classrooms and schools and skilful assessment to facilitate learning and reduce disparities,"(UNICEF, 2000).

Stakeholders in education have continued to decry overcrowded classrooms which they claim may have negative effects on learning outputs in public schools. The lack of equity in resourcing schools has led to quality of education being imbalanced across most African countries. Whereas quality of education is a broad aspect and may take a long discussion in trying to establish the various tenets of measurements, educationists have a consensus that examination scores are a viable proxy to assess the quality of education. The inputs in schools through a production function have been used in a lot of research work to be the influence of assessment outcomes.

In Kenya, examination scores are used to assess the amount of learning that takes place through the learning processes in schools. More than that, the scores are used in filtering pupils and students for transition to higher levels of education. Entry to secondary schools would require that pupils meet some minimum score. Transition into tertiary education would equally require a student to have met defined desirable grades. In the job market, employers place more weight on candidates who have higher examination achievements than those who perform comparatively lower.

1.3 Background

The Government of Kenya has made big and bold steps towards improving access to basic education since 2003. The efforts are reflected in the share of national resources that have been allocated to the education sector in the last thirteen years as well as a raft of policy measures that have been undertaken during the same period. These measures were taken to enhance access to education and ensure there is equity in the participation of learners in school. During the period 2000-2014 enrolment in primary education increased from 5.9 million to 9.9 million representing an increment of about 68%. The increment covered over age population that had not accessed school as well as population that annually got to school going age. More rapid growth in the enrolments was recorded in the rural areas. This may be attribute to the removal of school levies that were charged by schools prior to the period before 2003. The levies had been cited as serious hindrance to accessing basic education.

The primary Gross Enrolment Rate, a measure of education system coverage increased from 92.7% in 2000 to 115% in 2011 before improving to 103.5% in 2014. Overall, the figures indicate that the system has enough capacity to carry its populations school going age. There are however, regional disparities that still to be corrected by employing region specific programs to address lack of equity in access to education. Primary Gross Intake Rate (GIR) hit 102% in 2014 while the Primary Completion Rate (PCR) was 79.3%. The number of primary schools in the country have incredibly increased from 18,617 in 2000 to 29,460 in 2014 representing a growth of about 61%. The progress made in improving access to primary education is credited to successful implementation of key education programs including the Free Primary Education (FPE) which has boosted the enrolment rates; the School Feeding Program (SFP) which has enhanced retention in Arid and Semi-Arid Areas (ASAL); School Infrastructure which has addressed congestions in schools; and School Instructional Materials which eased the pressure of school materials acquisition from parents. There are a raft of other programs that the Government has implemented to advance education which may be obtained from the Ministry of Education Science and Technology (MOEST) as the custodian of education programs in the country.

Most of these programs were implemented in a 'one size fits all'fashion for instance the FPE assumed a capitation model where each pupil enrolled in a public primary schools was allocated an agreed per capita irrespective of their geographical background. The funds are disbursed to schools and are spent on teaching and instructional materials, school utilities and maintenance. It is important to note that during the conceptualization of the program, there was no distinction made between regions, whether they had any difference in terms of economic endowments and potentials when making the allocations. There was no variation made in the allocations to rural or urban pupils. An assumption was made that providing a flat rate capitation for each learner across the country would be sufficient in supporting education for all learners. With this support, learners are assumed to have been equalized across the country.

There is an acknowledgment of the achievements that have been registered this far by the education sector and specifically in primary education for the sake of this study. However, the achievements in increasing access to primary education notwithstanding, there is need to understand the story behind the programs implemented *vis-a-vis* the outcomes that pupils register at the end of each academic cycle. The country is endowed with diversity of geographical and other natural factors that may influence the education outcomes. Knowing that pupils vary in assessments not only because of their individual differences may help the country design better programs that will address disparity in education. In addition, knowing that a pupil in a school with a given set of characteristics is likely to perform in a certain predictable way should prompt the Government to adjust the inputs of the established model to establish an equitable system. This will also give parents the platform to make a choice of where to take their children and thereby instituting a competitive culture where schools hunger to satisfy their communities. Deliberate creation of varied models for each grouping or cluster may be desirous in optimizing the achievements from learning institutions. In deed the Government should have the motivation to measure the return on investments to education and to know where there could be better returns compared to other areas.

1.4 Statement of Problem

At the end of each academic year, learners who have been in schools for roughly eight years, having gone through instruction of the same curriculum, are subjected to a national examination. I say roughly eight years taking note that some pupils may stay in the primary cycle longer than eight years and this explains why the PCR in 2014 for instance was 79.3%. Adjusting for entry behaviour of children, the expectation from each of the individual would be minimal variation in the results of the national examinations. However, the scores from Kenyan National Examinations have depicted wide variation over the years. The variation has often been presented as pupil-pupil difference and noting the different means of different regions or administrative units.

There are grouping effects that may be responsible for some of the variations seen in pupils'scores. Pupils are grouped in classrooms; classrooms are grouped in schools; schools are grouped in districts; and districts are grouped in counties. These groupings have unique characteristics that may be responsible for some variation that is witnessed in the candidates'scores which seems to be ignored.

Application of appropriate models is required in effort to have the variance contribution of these groupings assessed and determined.

This far we have discussed the various areas where examination scores are applied in the life of a candidate with respect to their futures. If it is established that the variations of these scores are influenced by factors that cannot be controlled by the candidates there would be need to accord them the balancing effect they deserve. Of the 13 countries and economies that significantly improved their mathematics performance between 2003 and 2012, three also show improvements in equity in education during the same period, and another nine improved their performance while maintaining an already high level of equity proving that countries do not have to sacrifice high performance to achieve equity in education opportunities(OECD, 2013).

1.5 Objective of the Study

The main objective of the study is to develop an explanatory model to sufficiently explain the variations in pupils'scores. Specific Objectives The specific objectives of the study include to:

i. Establish whether grouping of pupils in schools have any effect on their examination scores

ii. Evaluate the proportion of variance explained by schools grouping effect

iii. Evaluate the suitability in the application of the model

1.6 Significance of the Study

(Vlaardingerbroek, Taylor and Heyneman, 2008) argue that, "Success in schooling system is one of the characteristics believed necessary for modern leadership. Although it is possible for leaders to emerge through experience, just good fortune, or military might, regardless, success in schooling is considered to be a sine qua non as an essential criterion of legitimacy". With this argument, a very heavy burden has been presented to educationists by the authors. First is to determine what success is in the context of education and subsequently define the measurements. The importance that countries place in the development of their education systems have been covered in the introduction section of this chapter. Having noted that, a very pertinent question is raised. If countries are to have a clique to pick modern leaders from then there is need to ensure that all children who are sent to schools 'succeed'. The chance of picking an effective leader according to Vlaardingerbroek *et al* is increased by reducing the variation between their successes.

When children join preparatory school, they are often told that they will be leaders of 'tomorrow'. As they graduate from the elementary level of the schooling system it is necessary that learners in whose hands the leadership of 'tomorrow'demonstrate acceptable success. If the success of learners is to be measured by what they score from school then there is need to ensure that education systems provide the children with environments that can facilitate their success. Education systems must strive to create an equal playing field for learners so that the variations that may be exhibited from assessments and examinations may be down to individual capacities.

Showing that schools account for significant variation on the learners scores will be a big impetus for education planners to review education production functions to establish the optimal level of providing resources to schools for the development of a sustainable and equitable system. As (Gustafsson, 2007) notes, appreciation of the inputs in education programs relative to the outputs becomes very important. I have noted that there is not much application of the mixed model in the country in the analysis of assessments and examinations results. This study seeks to add to the existing analyses and I believe it will spur the national application in the analysis of examination scores for consideration of equitable transition to secondary schools.

Chapter 2

Literature Review

2.1 Introduction

This chapter presents the background of examination systems and a review of existing knowledge from papers and texts published or documented by researchers in education and other related fields. The review provides insight to approaches that existing work has employed in analyzing education outcomes especially outcomes from examination and assessments. The chapter also discusses the theories of the widely used methodologies in education research; their merits as well as demerits. The chapter finally discusses the model selected for this research setting the basis for its preference.

2.2 Education Assessment Systems

(Braun and Kanjee, 2006) opine that "although assessment is often seen as a tool to measure the progress of individual students, it also allows individuals, communities, and countries to track the quality of schools and educational systems". It is true that assessments help in examining the achievement levels of the system. From assessment scores, the system accounts for what pupils and/or students learn. It is sort of a feedback expressing the process of transfer of knowledge. If the process is effective then it is expected that learners would exhibit high scores in the assessment. This would amount to assessing the state of the quality of education systems. In strategic planning, corporates often motivate themselves that 'what gets measured gets done'. I believe education assessments equally operate in the same spirit.

In Kenya assessment systems have been used for nearly the same purposes. Examinations and assessments have been used in assessing the knowledge and competency that learners get out of any considered cycle of education for instance Early Grade Reading assessment was mounted to establish the reading skills of learners in mother tongue, English and Kiswahili (Piper, 2010); The principal purpose of SACMEQ is to track the learning competencies achieved by learners at primary school level (Wasanga, Ogle and Wambua, 2010); and the results of Uwezo show that school going children are not achieving desirable competencies in literacy and numeracy (Uwezo Kenya, 2013). All these point out some of the reasons for which education assessments have been carried out in the country.

The national end of education cycle examinations results have been used to filter learners who are deemed 'ready'for the succeeding level of education. The Kenya Certificate of Primary Education Examination results for instance are used in establishing those who transit to secondary education while the Kenya Certificate of Secondary Examination results are used in placing learners into middle level colleges and institutions of higher learning according to the standards they achieve. The standards are set by the MOEST through public policies. In some cases assessments have been carried out for ad hoc purposes, may be to get a feeling of what is happening in the schooling system. "In theory, if policymakers have access to reliable information on educational quality in specific schools and make this information available to the aware public, then students and parents may be better able to choose among educational options and demand education of higher quality,"(Braun and Kanjee, 2006). This is a theory that should have a practical place in the 21st century if education has to make meaning to those who seek it as well as to those who invest in it especially governments, communities and private entities.

2.3 Previous Studies on Education Assessments

As earlier mentioned, all learners in countries that have nationally regulated education systems are subjected to similar curricula. In Kenya, all learners are subjected to the same curriculum (8-4-4) which is regulated nationally by the Ministry of Education Science and Technology. Assessment systems have over the years exhibited that despite this uniformity of curriculum, there exists variation in the assessment outcomes. Most of the education research work done in the country have used the multiple regression model in the analysis.

In a study to establish the influence that final examinations in primary schools have in the final secondary examinations score of students from Nyamira District, Nyamira County, Ondima *et al* used the linear regression to demonstrate their idea. "Regression analysis will be used to analyse data which will be useful in measuring entry scores in KCPE and final performance in KCSE. Given the reliability of regression analysis of the data, the results could be used to alert the management, stakeholders and parents, the level of learners and the strategies needed for them to achieve better results,"(Ondima *et al*, 2013). It is important to note that the students who were selected for the study were nested in schools and the schools are further nested to administrative zones.

The linear regression in this case has not taken into account that students grouped in the same school may have similar achievement tendencies that may arise from their learning together. In any case the number of students selected from each school was not uniform thus making the choice for standard regression inappropriate for their study because of the unbalanced structure of the data used in the analysis. In their conclusion, Ondima *et al*, found out that KCPE score for each student included in the study did not significantly influence the KCSE scores for the same students. Since the students'scores four years earlier were tacked and their compared to their KCSE performance, the design of the study ought to be that of a repeated measures. In case the design could be modified to have the same number of students drawn from each school, repeated measures anova could have been applied. The effect of grouping in schools was not taken into account either. Perhaps a complex situation was how to handle students in a given school considering that they attended different schools for the KCPE assessments. As it was, the mixed model should have sufficed.

(OECD, 2013) reports on the Programme for International Student Assessment (PISA) which is undertaken by sixty five OECD countries to assess “the extent to which students near the end of compulsory education have acquired some of the knowledge and skills that are essential for full participation in modern society, particularly in mathematics, reading and science”. PISA “gathers information from students about their learning environment, educational experiences, and attitudes towards education,”, “Analyses of PISA data provide information on the relative performance of students and on the differences between student environments, attitudes, and experiences within and across countries,”(Kastberg, Roey, Lemanski, Chan and Murray, 2014).

With the design of national and international variables built in the assessment tools, it becomes very easy to establish the within country and between country variation and of course the proportion of variation that this kind of data structure can explain. The Rasch model, was used in scaling the learners'data to provide insight into the performances which indicated high disparities between countries and well as gender imbalances. The program's results are used by policy makers from participating countries to develop their respective education systems.

The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ), another system of assessment, carries out large-scale cross-national research studies in member countries in the Southern and Eastern Africa region. The aim of the assessments carried out by the consortium is to evaluate the status of schools and learners'achievements in literacy and numeracy, (ACER, 2015). Over the years there has been an attempt to assess the achievements in literacy and numeracy with a view to informing policy makers on reviewing what works and what does not work in the schooling systems. The SACMEQ has implemented four series of assessments since its inception in the 1990's, the latest being SACMEQIV which was conducted in 2014.

In the SACMEQ III results, linear regression was used to establish the significance of selected predictor variables on the assessment scores of grade six learners (Wasanga et al, 2010). The predictors ranged from learners to school and teacher variables. The account given by the procedure used by SACMEQ demonstrates the first stage of fitting a mixed effects model where all the independent variables are treated as fixed effects where all explanatory variables are treated as fixed and the OLS procedure is used in estimating the parameters. The researchers found 70 variables in the study to be significant at (α =0.05), a finding that should have triggered tests of misspecification of the model as the variables may be too many. To account for grouping effects in a study, subsequent steps have to be undertaken and decision has to be made on which variables are to be made fixed and which ones are to be considered as having random effects to learners'achievements in reading or mathematics. (Wasanga *et al.*, 2010) reckon that multilevel modeling was conducted at two levels i.e. pupil level and school level.

The researchers have failed one, to demonstrate the procedure which was used in the second level of analysis. Readers are left to wallow in dilemma of not knowing which pupil level or schools level variables were treated as either fixed or random. The results provided for the SACMEQ III project in the case of Kenya did not indicate the intercepts for the random variables included in the fitted model. The variance accounted for by the grouping of pupils in a school is equally not given which raises questions on the comprehensiveness of the procedure and methodology. It is evident that there is appreciation of the mixed effects model, or multi-level modeling as the authors have referred to it, in analysis of multi-level data is there. However, the SACMEQ III project has not demonstrated to the latter the application of the methodology.

A similar case is presented during the development of the multi-level modeling where a study had been conducted on primary schools learners in the 1970's. The study looked at formal styles of teaching reading and the progress made by such learners as opposed to learners who were not exposed to such styles. Analysis was carried out using standard regression analysis treating learners as the only unit of analysis. The analysis did not put any emphasis on the effects of grouping learners into teachers grouping (Goldstein, 1999). In later work the same data were subjected to multi-level analysis which then showed that some of the variables had lost their significance to the grouping effects. This fact raises a very fundamental question when dealing with multi-level data.

In his working paper on the development of an equitable public schooling system in South Africa, (Gustafsson, 2007) used hierarchical linear modeling to establish the effect grouping pupils in schools had on their learning outcomes. He established that 55% of the total variation was due to schools effect. Learners accounted for only 10% of the variation. This points to a systemic neglect rather than learners'inability to succeed in their education endeavours.

2.4 Summary of Findings

While I note the incredible work that has been carried out by researchers in the field of education, I have also noted that there is a wide usage of trivial models to explain observed phenomena. It is with this understanding that I propose the application of linear mixed effects model in the analysis of learners'outcomes. This will help in developing adequate program that address disparities in education.

In the next section, I have provided a review of the linear regression modeling detailing its application as well as the limitations. This way, there is an understanding of the basis upon which my critiques are based. The transition into hierarchical modeling is also made easier for my future readers of this work.

2.5 Linear Regression Model

To provide a transition from the most used model in social research to the model that has gained prominence in the past four decades, the standard regression model is reviewed in this section. Some aspects of the standard linear regression remain very useful in the mixed effects model, especially in the estimation of fixed effects parameters. The most basic model that may be used to explain the relationship between variables, one being continuous and the other may take any form, is the linear model. The direction and magnitude of the association may be deduced simply by employing the simple linear regression model.

“Simple linear regression is the most commonly used technique for determining how one variable of interest (the response variable) is affected by changes in another variable (the explanatory variable). The terms ”response” and ”explanatory" mean the same thing as "dependent" and "independent", but the former terminology is preferred because the "independent" variable may actually be interdependent with many other variables as well,"(Kirchner, 2001).

"In spite of the availability of highly innovative tools in statistics, the main tool of the applied statistician remains the linear model. The linear model involves the simplest and seemingly most restrictive statistical properties: independence, normality, constancy of variance, and linearity. However, the model and the statistical methods associated with it are surprisingly versatile and robust. More importantly, mastery of the linear model is a prerequisite to work with advanced statistical tools because most advanced tools are generalizations of the linear model,"(Rencher and Schaalje, 2008). The general form of the linear regression model is given as:

$$y = X\beta + \varepsilon \tag{2.1}$$

where y is an $n \times 1$ vector of observed responses, X is an $n \times p$ matrix of constants, β is a $p \times 1$ vector of unknown parameters, and is an $n \times 1$ vector of random errors. The reason the model represented in equation (2.1) is called linear comes from the relationship between the observed response vector y and the parameter β. The mean of the observed y values is linear in β. The simple form of the linear regression is considered when there is just one independent variable explaining the observations on the response variable.

$$y_i = \beta_0 + \beta_1 X + \varepsilon_i \tag{2.2}$$

The parameters in equation (2.2) are estimated using Ordinary Least Squares (OLS) method where the squared residuals obtained from the fitted model are minimized i.e.

$$\varepsilon_i = y_i - \hat{y} \tag{2.3}$$

$\hat{y}$ is the estimated value of the response variable using the line of best fit while $y_i$is the observed response variable. Since summing individual residuals will obviously result to zero, they are squared such that:

$$\Sigma\varepsilon_i^2 = \Sigma(y_i - \hat{y})^2$$

But

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_1$$

$$\Sigma\varepsilon_i^2 = \Sigma(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1))^2 \tag{2.4}$$

To solve for the parameters and we obtain partial derivatives of the sum of the squared residuals and minimize them:

$$\frac{\delta\Sigma\varepsilon_i^2}{\delta\beta_0} = 2\Sigma(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1)) \times (-1) \tag{2.5}$$

$$\frac{\delta\Sigma\varepsilon_i^2}{\delta\beta_1} = 2\Sigma(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1)) \times (X_1) \tag{2.6}$$

Solving equation (2.5) and (2.6) simultaneously, we obtain

$$\beta_1 = n\Sigma X_i Y_i - \frac{\Sigma X_i Y_i}{n\Sigma X_i^2 - (\Sigma X_i)^2}$$

$$\beta_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

We note that the linear model described in equation (2.2) allows for only one predictor. However, in the real world most of the observed phenomena in day to day situations are composites of multiple factors. We may take the example of a political election where aspirants seek votes from the electorate. There are several factors that may influence voters to give their vote to a given aspirant and not the next aspirant. These factors may include the aspirant's age, sex, marital status, religious affiliation, and maybe perceived or real economic endowment. In this case, the general form of the linear model described by equation (2.2) may be specified to allow for more than one causal variable. In the example of a political aspirant, all the five listed variables will be included in the resultant model. In this case the Linear Regression Model changes from Simple to Multiple Linear Regression Model which may be represented by equation (2.7) and the normal equations used in the estimation of model parameters can be developed as in equation (2.9), (2.10) and (2.11).

$$y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik} + \varepsilon_i \tag{2.7}$$

$$\Sigma\varepsilon_i^2 = \Sigma(y_i - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik}))^2 = Q \tag{2.8}$$

$$\frac{\delta Q}{\delta \beta_0} = 2\Sigma(y_i - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik})) \times (-1) \tag{2.9}$$

$$\frac{\delta Q}{\delta \beta_1} = 2\Sigma(y_i - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik})) \times -(X_{i1}) \tag{2.10}$$

$$\frac{\delta Q}{\delta \beta_k} = 2\Sigma(y_i - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + ... + \beta_k X_{ik})) \times -(X_{ik}) \tag{2.11}$$

These sets of normal equations in equations (2.9), (2.10) and (2.11)can be represented in a matrix form as:

Table 2.1: Normal Equations for Estimating Multivariate Linear Regression

	y	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	...	$\hat{\beta}_k$
1	Σy_i	$n\hat{\beta}_0$	$\hat{\beta}_1 \Sigma X_{i1}$	$\hat{\beta}_2 \Sigma X_{i2}$	...	$\hat{\beta}_k \Sigma X_{ik}$
X_{i1}	$\Sigma X_{i1} y_i$	$\hat{\beta}_0 \Sigma X_{i1}$	$\hat{\beta}_1 \Sigma X_{i1}^2$	$\hat{\beta}_2 \Sigma X_{i1} X_{i2}$	...	$\hat{\beta}_k \Sigma X_{i1} X_{ik}$
X_{ik}	$\Sigma X_{ik} y_i$	$\hat{\beta}_0 \Sigma X_{ik}$	$\hat{\beta}_1 \Sigma X_{i1} X_{ik}$	$\hat{\beta}_2 \Sigma X_{i2} X_{ik}$	...	$\hat{\beta}_k \Sigma X_{ik}^2$

The multiple linear regression in Table (2.1) above is of the form $y = X\beta + \varepsilon$ and its parameters are estimated by:

$$\underline{\hat{\beta}} = (X^{'}X)^{-1}X^{-1}\underline{y} \tag{2.12}$$

Assumptions and Estimation of Linear Regression Model

Application of linear regression models makes the following fundamental assumptions:

i. The data used in analysis is representative of the population under study

ii. The relationship between the dependent and independent variables is linear.

The parameters can be estimated using the Ordinary Least Squares (OLS) approach or the method of maximum likelihood. In estimating the parameters of the equation each of the residuals for corresponding observations are assumed to be normally and independently distributed with mean zero and variance of δ. The variance of the residuals is assumed to be constant. This assumption together with the assumption on representativeness and linearity must be met if the parameters estimated using the model are to be unbiased estimators of the population parameters. Apart from estimation of the population parameters, these assumptions are necessary in testing hypotheses and constructing confidence intervals within which the decision taken on hypotheses testing hold.

2.5.1 Limitations of the Linear Regression Model

Inasmuch as the linear regression model has been widely used to analyse numerous real world phenomena there are some challenges it did not address which motivated the development of further models. "Prior to the development of HLM, hierarchical data was commonly assessed using fixed parameter simple linear regression techniques; however, these techniques were insufficient for such analyses due to their neglect of the shared variance,"(Woltman, Feldstain, Mackay and Rocchiet, 2002).

The linear regression model assumes a single mean with no regard for additional intercepts that may be realized due to grouping. In their demonstration of random variation concept, (Pinheiro and Bates, 2000) established that introducing the grouping effect of Rails, "there is considerable variation in the estimated mean travel times per rail. The residual standard error obtained for the fixed-effects model δ =4.0208, is about one-sixth of the corresponding estimate obtained for the single-mean model". Other weaknesses include the assumption that an observation on the dependent variable is due to an independent variable. In some cases, the conclusion of causality may just be mere coincidence in which case there could be other variables excluded from the model or the phenomena may have some other external factors influencing the observations.

2.6 Hierarchical Linear Modeling

"Hierarchical Linear Modeling (HLM) is a complex form of Ordinary Least Squares (OLS) regression that is used to analyze variance in the outcome variable when the predictor variables are at varying hierarchical levels,"(Woltman et al, 2002). In contrast to the linear model which assumes that all independent variables influence the dependent variable from the same level and therefor ignores shared variance in the grouped subjects, hierarchical modeling takes into account the effects of the groups into which the subjects are. Thus the linear mixed effects model is applicable in the analysis of grouped data.

"In a mixed model, the total residual variation of the observations is divided to within-group and between-group variation. However, after estimating these variance components, we can use the observations of our data to predict also effects for individual groups,"(Mehttalo, 2013). "A grouped dataset may have either a single level of grouping or multiple levels of grouping,"(Mehttalo, 2013). In the single case we may consider pupils grouped in a classroom such that we would be interested to know whether grouping the pupils in a classroom explains any proportion of the variability from the next classroom. In a two level case we would consider pupils grouped into classrooms and the classrooms are further grouped in schools. Consequently, the hierarchical model is used to establish the proportion of variance accounted for by both classroom and school effects compared to the proportion accounted for by individual pupil difference.

2.6.1 Crossed vs Nested Data Structure

The structure of data in multi-level modelling may be crossed or nested. "In a two-way design, the analysis is considered crossed if each level from one way is contained in each level of the other way. In this design, every person (unit of analysis) has a score in every cell,"(Kyle, 2002). In an education context we would consider the example of classrooms and teachers. Teacher x would be available to teach in classroom A and classroom B. This design would be crossed since for each of the classes assessed, teacher x has a score. In a nested case, there is seamless hierarchy where no interaction exist between pupils in a given school and the next school. The uniqueness of a school is limited to the pupils who are grouped into it. If higher hierarchies are available, the seamlessness will be sustained, there will be no interaction between schools in a given location for instance.

2.6.2 Balanced vs Unbalanced Data Structure

In the analysis of variance where treatments effects on experimental units are assessed, balance refers to allocation of equal number of experimental units to each of the treatments to be evaluated. In the unbalance case, the number of subjects is not same for the grouping levels considered. In the case of this study, the number of pupils who are considered per school is not the same.

2.6.3 Fixed vs Random Effects

The notion of fixed and random effects variables can be so challenging to apply to a dataset as there are no hard and fast rules to what one would consider a fixed or a random variable. However, based on sampling, "A fixed effect factor is a factor whose levels are the only possible levels in the population being studied. This is opposed to a random effect factor whose levels in the study are just a sample of all the other possible choices,"(Onyango, 2009). "A group effect is random if we can think of the levels we observe in that group to be samples from a larger population,"(Taylor, n.d.).

(Winter, 2006) attributes a fixed variable to one that is measured with an absolute precision. Usually, it is assumed that the values of a fixed variable remains the same across studies such that the value assumed by a variable in study a will be similar to the value assumed by the same variable in study b. 'Random variables'are assumed to be values that are drawn from a larger population of values and thus will represent them. With this understanding and subsequent application to the pupils'score dataset, sex of a pupil which is measured as either male or female, is considered to be a fixed effect in my model. In the case of schools, since the schools selected for analysis are due to some systematic or non-probabilistic chances, the effects they have on the model will be random.

2.6.4 Inter Class Correlation

One of the main reasons the mixed model is preferred is its ability to estimate the effects of a higher level hierarchy on lower level subjects that are grouped therein. One of the statistics that we look out for is the inter-class-correlation, in the case of this study inter-school-correlation. From the estimation of the variance components of the model we have the total variance due to pupil and schools given as:

$$var(y|\underline{\beta}, X_{ij}) = var(\upsilon_{0j}, \varepsilon_i j) = {\delta_{u0}}^2 + {\delta_{\varepsilon 0}}^2$$

Where ${\delta_{u0}}^2$ is the school variance while ${\delta_{\varepsilon 0}}^2$ is the pupil variance.

The inter-class-correlation is given by:

$$\rho = \frac{{\delta_{u0}}^2}{{\delta_{u0}}^2 + {\delta_{\varepsilon 0}}^2}$$

The inter-class-correlation estimates the proportion of variance due to the grouping effect (Goldstein, 1999). When the inter-class-correlation is lower in multilevel model, there the estimates are reasonably close to those obtained by the standard OLS estimates,(Goldstein, 1999). This fact can be used as a test to validate the choice for a mixed model instead of standard regression model. The variation between pupil to pupil, the within school variation, is given by:

$$cov(\upsilon_{0j} + \varepsilon_{i_1 j}, \upsilon_{0j} + \varepsilon_{i_2 j}) = cov(\upsilon_{0j}) = {\delta_{u0}}^2$$

2.7 Non Normal Residuals and Non Continuous Response Variables

In the real world, it is obvious that not all phenomena will have observations record-able in continuous forms. The other challenges that may arise are the violation of critical assumptions like normality, homoscedasticity. There are other models that will come in handy in dealing with this dilemma. "Generalized linear mixed models provide a means of modeling these deviations from the usual linear mixed model,"(Kachman, 2000). Just like in linear models, generalized linear models come into play where assumptions in the former are violated, the generalized linear mixed models have been widely used in the same context for the linear mixed effects models. The focus remains on the link function. However, on the generalized linear mixed effects model the focus is on the inverse of the link function.

Chapter 3

Methodology

3.1 Introduction

This chapter gives the functional details of the model described in chapter two. It discusses the data structure as used in the research project; formulation of the linear mixed model; the structure of the model and its components; the basis and steps for model selection; model assumptions and tests of assumptions of the model; estimation of the model parameters; hypothesis testing and confidence intervals; and prediction using the fitted model.

3.2 Data Structure

The data used in this research was obtained from the results of the 2013 Kenya Certificate of Primary Education examinations which are administered by the Kenya National Examinations Council (KNEC). The research has made a deliberate focus on Homa Bay County which has six administrative districts and over one thousand schools. We note that not all candidates who sat for examination in 2013 have been considered in this study as some of them sat for the examinations in private centers which are not schools. Since the study aims at predicting future outcomes of a pupil's final examination score based on individual characteristics of a pupil who sat for an exam as well as the characteristics of the school the pupil took the exam from, private examination centers had to be omitted. The rationale being that their characteristics are not available.

The study has adopted a two level hierarchical model. Pupils and their individual characteristics have been considered in the first level and are nested in the schools from which they took their exams, considered as level two. We wish to analyze the effects of independent school characteristics, level two variables, on the pupils'score, a level one variable.

In multilevel modeling, the dependent variable has to be a level one variable thus a Pupil's Kenya Certificate of Primary Education Score has been treated as that. Explanatory variables may assume either levels of the hierarchy for instance Age and sex are explanatory variables on level one of the hierarchy while schools status and location are on level two. Figure 3.1 shows a sample of the data used in this research work.

Age The age of the pupil at the time of the examination
Sex The biological category of the pupil, male or female
Status The ownership of the school, whether private or public
Location The location of the school, either rural or urban.

3.3 The Linear Mixed Effects Model

"Mixed-effects models are primarily used to describe relationships between a response variable and some covariates in data that are grouped according to one or more classification factors,"(Pinheiro and Bates, 2000). Unlike the linear model,

Pupil	Sex	Age	KCPE_Score	School	Status	Location
689	Male	14	293	25	Public	Rural
1241	Male	18	288	46	Public	Rural
1294	Male	16	213	47	Public	Rural
1416	Male	17	NA	53	Public	Rural
1548	Male	15	274	57	Public	Urban
2064	Male	17	260	73	Public	Rural
4269	Male	12	330	162	Public	Urban
6300	Male	17	289	246	Public	Rural
6469	Male	17	276	252	Public	Rural
6667	Male	16	248	260	Public	Rural
8032	Female	15	197	312	Public	Rural
8358	Male	16	258	321	Public	Urban
10023	Male	15	315	387	Public	Rural
10663	Female	17	279	409	Public	Urban
10711	Male	17	185	410	Private	Rural
11897	Female	15	196	456	Public	Rural
13861	Male	14	366	530	Public	Rural
14142	Male	14	274	540	Public	Rural
15824	Male	14	220	609	Public	Rural
16121	Male	13	350	620	Public	Rural
16173	Female	15	272	620	Public	Rural
16183	Male	15	267	620	Public	Rural
18523	Female	14	168	701	Public	Urban
19543	Male	17	150	738	Public	Rural
21094	Male	13	339	806	Public	Rural

Figure 3.1: **Sample of Pupils' Score Data**

the mixed effects model has an additional random component which linearly influences the response variable. The random component of the mixed effects model partitions the error term in the linear model thus reducing the proportion of unexplained variability in the resultant model.

3.3.1 Model Formulation

We begin from the standard linear regression. A pupil's score is given by:

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{3.1}$$

For the multilevel model we consider the modification of the standard linear regression equation such that we have

$$y_{ij} = \beta_{0j} + \beta_{1j} Age_{ij} + \beta_{2j} Sex_{ij} + \varepsilon_{ij} \tag{3.2}$$

y_{ij} is the score of the ith pupil in the jth school for i=1,2,3,,N and j=1,2,3,M; β_0 is the regression constant; β_{1j} is the regression slope; and ε_{ij} is the residual.

From (3.2) each level two unit, schools in this case, has its unique level one regression constant (β_{0j}). We also note that each school has a unique pupil level regression slopes based on sec and age. These constants and regression slopes vary from school to school. In order to sufficiently account for the effects of school level variables, we consider parameters of equation (3.2) and use them to develop school level equations. Each of the coefficients in (3.2) will yield a level two equation given in (3.3), (3.4) and (3.5).

$$\beta_0 j = \gamma_{00} + \gamma_{01} SchoolStatus + \gamma_{02} SchoolLocation + \upsilon_{0j} \tag{3.3}$$

$$\beta_1 j = \gamma_{10} + \gamma_{11} SchoolStatus + \gamma_{12} SchoolLocation + \upsilon_{1j} \tag{3.4}$$

$$\beta_2 j = \gamma_{20} + \gamma_{21} SchoolStatus + \gamma_{22} SchoolLocation + \upsilon_{2j} \tag{3.5}$$

Here, $\gamma's$ are the school level regression coefficients and do not vary from school to school. For instance the coefficients for a rural school remains the same across all rural schools in the sample. In fact, it is for this reason that the regression coefficients are modeled without subscripts. υ_{0j},υ_{1j} and υ_{2j} are random variations between schools.

$$E\begin{bmatrix}\upsilon_{0j}\\ \upsilon_{1j}\\ \upsilon_{2j}\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}; var\begin{bmatrix}\upsilon_{0j}\\ \upsilon_{1j}\\ \upsilon_{2j}\end{bmatrix} = \begin{bmatrix}\tau_{00} & \tau_{01} & \tau_{02}\\ \tau_{10} & \tau_{11} & \tau_{12}\\ \tau_{20} & \tau_{21} & \tau_{22}\end{bmatrix}$$

$var(\varepsilon_{ij}) = \delta^2; \quad var(\upsilon_{0j}) = \tau_{00}; \quad var(\upsilon_{1j}) = \tau_{11}; \quad var(\upsilon_{2j}) = \tau_{22}$

$cov(\varepsilon_{ij}, \upsilon_j) = 0$

Substituting equations (3.3), (3.4) and (3.5) into (3.2) we have

$$y_{ij} = \gamma_{00} + \gamma_{01}SchoolStatus + \gamma_{02}SchoolLocation + \upsilon_{0j}$$

$$+Age(\gamma_{10} + \gamma_{11}SchoolStatus + \gamma_{12}SchoolLocation + \upsilon_{ij})$$

$$+Sex(\gamma_{20} + \gamma_{21}SchoolStatus + \gamma_{22}SchoolLocation + \upsilon_{2j}) + \varepsilon_{ij}$$

Which can be simplified into

$$y_{ij} = \gamma_{00} + \gamma_{01}Age + \gamma_{20}Sex + \gamma_{01}SchoolStatus + \gamma_{02}SchoolLocation +$$

$$\gamma_{11}SchoolStatus * Age + \gamma_{12}SchoolLocation * Age + \gamma_{21}SchoolStatus * Age +$$

$$\gamma_{22}SchoolLocation * Sex + \upsilon_{0j} + \upsilon_{1j}Age + \upsilon_{2j}Sex + \epsilon_{ij}$$

Notice that the resultant equation has a grand mean, level one and level two predictors which in some cases are interacting and finally the variance component which is composed of school level randomness and the pupil level variation.

The model formulated can be represented in matrix form as:

$$y = X\beta + Z\upsilon + \varepsilon \tag{3.6}$$

y is the examination score for each of the pupil in the study. X and Z are $n \times p$ and $n \times q$ design matrices respectively which give the relationship between the fixed effects and random effects covariates and the observations. β is an unknown vector of fixed effects. υ is also an unknown vector of random effects which is normally distributed with mean and variance-covariance matrix $var(\upsilon) = G$. Finally e is an unknown $n \times 1$ matrix of random errors which again are assumed to be normally distributed with mean $E(\varepsilon) = 0$ and variance $var(e) = R = \delta_{\varepsilon}^2 I$. The identity matrix in the variance of the residual is of size $n \times n$.

$$E(\varepsilon) = 0; E(\upsilon) = 0$$

$$\varepsilon \sim N(0, R);\ where\, R = \delta_{\varepsilon}^2 I$$

$$\upsilon \sim N(0, G);\ where\, G = \delta_{\upsilon}^2$$

The expectation of the formulated model is given by:

$$E(y) = E(X\beta) + E(Z\upsilon) + E(\varepsilon)$$

$$= \beta E(X) + ZE(\upsilon) + E(\varepsilon)$$

$$= X\beta$$

We note that the variance component in a mixed effects model is composed of the variance of the subjects as well as the group variance. Assuming that u and e suffer no multi co-linearity, the covariance matrix associated with the response variables is given by:

$$var(y) = var(\varepsilon) = var(Z\upsilon + \varepsilon) = ZGZ^T + R$$

The model described this far requires that we satisfy that the data to be applied satisfy the model. We take a pause in the theory of the model to cover model selection. The assumptions of the model and estimation of model parameters as well as tests of hypotheses are covered in subsequent sections.

3.3.2 Model Selection

In order to select the best model that suits the pupils'score data, iterative models have been fitted manipulating the variables in the data, making them fixed and random and assessing their effects when applied on the model defined in equation (3.17). In all models fitted, Age and Sex of a pupil are considered as fixed effects variables. The first model (Model1) considers alongside Age and Sex, Location and School Status as fixed effects predictors. Both possible levels of location (rural and urban); and both levels of school status (private and public) have been selected leaving no chance for randomness hence the treatment as fixed effects. Since Age and Sex are level one variables, an interaction with level two variables is considered. Schools are treated as random effects as not all the schools in the county were selected for this study. Inasmuch as school location is a fixed effect in the model, schools are located either in the rural or urban area and the model has been fitted with this consideration the Restricted Maximum Likelihood method.

$Model1 = lmer(KCPEScore\ Age * Location + Age * Status + Sex * Location + Sex * Status + (Location|School)$

where (*) represents full interaction between variables

The second model has dropped the interaction between age and school status and location so that the fitted model is given by:

$Model2 = lmer(KCPEScore\ Age * Status + Sex * Location + Sex * Status + (Location|School)$

Iteratively the following models have been fitted to increase the range of choices from which the eventual model is selected.

$Model3 = lmer(KCPEScore\ Age + Sex * Location + Sex * Status + (Location|School)$

$Model4 = lmer(KCPEScore\ Age + Sex * Status + Location + (Location|School))$

$Model5 = lmer(KCPEScore\ Age + Sex + Status + Location + (LocationDistrict))$

$Model6 = lmer(KCPEScore\ Age + Sex + Status + Location + (1|School)$

There are more permutations of models that could be fitted but for the time being we concentrate on the six and evaluate the best to be selected. It is noted that there exist several criteria that may be used to select suitable models. One of the criteria is the R2 which tends to be better as the model grows bigger. In effect models may end up having too many variables in an effort to maximize R2. The other criteria is the adjusted R2 whose values decrease with increasing size of the model. The two are largely used in the standard linear regression models and caution should be taken to test against model misspecification when using them as criteria for model selection.

Misspecification tests that include the Regression Specification Error Test (RESET) can be carried out to check for any omission of critical variables or inclusion of irrelevant variables in the model being tested. Other criteria include the Akaike Information and Bayesian Information Criteria which offer the relative estimate of information lost when a given model is used to represent the mathematical processes and interactions that yield the specified response variable. The criteria examines the goodness of fit of the model in question and balances that with its complexity. All the models have been fitted using the Restricted Maximum Likelihood Method. To obtain the AIC, analysis of variance procedure has been conducted in R comparing all the fitted models. The results from the analysis of the variance is shown in Table (3.1) below.

The output results include the Akaike Information Criteria (AIC) as well as Bayesian Information Criteria (BIC) which are computed as:

$$AIC = -2logLik + 2k$$

$$BIC = -2logLik + klog(N)$$

Table 3.1: **ANOVA Output for Model Comparison**

Model	Df	AIC	BIC	LogLik	Deviance	Chisq	ChiDf	Pr(Chisq)
Model6	5	222619.4	222659.3	-111304.7	222609.4	NA	NA	NA
model5	7	222612.5	222668.3	-111299.2	222598.5	10.9635	2	0.00416
model4	8	222599.2	222662.9	-111291.6	222583.2	15.2854	1	9.2424e-05
model3	9	222449.5	222521.2	-111215.7	222431.5	4.2566	1	0.03909
model2	11	222443.3	222531.0	-111210.6	222421.3	10.2122	2	0.00605
model1	13	222438.0	222541.6	-111206.0	222412.0	9.3235	2	0.00944

Here k is the number of parameters in the considered model while N is the total number of observations in the model. Under these definitions, smaller is better. That is, if we are using AIC to compare two or more models for the same data, we prefer the model with the lowest AIC, (Pinheiro and Bates, 2000). The Bayesian criteria employs similar evaluation of smaller is better although its values will obviously turn out larger than the AIC for large samples. Using Akaike Information Criteria, model1 is selected as it has the lowest value and will be used in subsequent sections for analysis and discussion.We recall that the model is given by:

$Model1 = lmer(KCPEScore\ Age * Location + Age * Status + Sex * Location + Sex * Status + (Location|School)$

3.3.3 Parameter Estimation

From the formulation of the mixed model, it is clear that there exist the observed part of the model and another that are unknown. Y,X and Z are observed. β,υ,R,and G are unknown. The unknown parameters are estimated in two folds, the () are estimated using the Best Linear Unbiased Estimator (BLUE) while the vector of the random effects (υ) are estimated using the Best Linear Unbiased Predictors (BLUP) method. We note that the (β) are fixed parameters and estimation methods can be used absolutely. The (u) on the other hand are not fixed occurrence and are due to chance. "estimation of a random quantity is often called prediction to emphasize the fact we are trying to get our hands on something that is not fixed and immutable, but something whose value arises in a random fashion,"(Davidian, n.d.) We begin by estimating the variance components. The results of which are used in estimating the fixed and random effects parameters.

Estimation of Gkelihood method of estimation is used in estimating the variance components (covariance matrices G and R) of the random effects parameters. "The maximum likelihood procedure produces biased estimates of the random parameters because it takes no account of the sampling variation of the fixed parameters. This may be important in small samples, and we can produce unbiased estimates by using a modification known as restricted maximum likelihood (REML),"(Goldstein, 1999) . The method is based on the marginal model.

$y = X\beta + \varepsilon^*$; $\quad \varepsilon^* = Z\upsilon + \varepsilon$ **such that** $\varepsilon^* \sim N_n(0, V)$; $\quad V = ZGZ^T + R$

Assuming that the variance components G and R have some parameter α, then

$$V(\alpha) = ZG(\alpha)Z^T + R(\alpha)$$

The marginal log likelihood of the model is thus given by

$$l_{(R(\alpha))} = In(\int(\beta, \alpha)d\beta)$$

$$\int(\beta, \alpha)d\beta = \int \frac{1}{(2\pi)^{(n/2)}}|V(\alpha)|^{(-1/2)} + exp\{\frac{1}{2}(y - X\beta)^t V(\alpha)^{-1}(y - X\beta)\}d\beta$$

But

$$(y - X\beta)^t V(\alpha)^{-1}(y - X\beta) = \beta^t X^t V(\alpha)^{-1}X\beta - 2y^t V(\alpha)^{-1}X\beta + y^t V(\alpha)^{-1}y$$

$$= (\beta - B(\alpha)y)^t A(\alpha)(\beta - B(\alpha)y) + y^t V(\alpha)^{-1} - y^t B(\alpha)^t A(\alpha)B(\alpha)y$$

With this, and R

Restricted Maximum Li

$$\int(\beta,\alpha)d\beta = \int\{\frac{|V(\alpha)|^{-1/2}}{(2\pi)^{n/2}}exp\{\frac{1}{2}(y^t[V(\alpha)^{-1} + B(\alpha)^t A(\alpha)B(\alpha)]y)\}d\beta\times$$

$$\int exp\{\frac{-1}{2}(\beta - B(\alpha)y)^t A(\alpha)(\beta - B(\alpha)y)\}d\beta$$

Here we note that

$$B(\alpha) = A(\alpha)^{-1}X^t V(\alpha)^{-1} \quad and \quad A(\alpha) = X^t V(\alpha)^{-1}X$$

We also note that

$$\hat{\beta} = (X^t V^{-1}(\alpha)X)^{-1}X^t V(\alpha)^{-1}y$$

$$= A(\alpha)^{-1}X^t V(\alpha)^{-1}y$$

$$= B(\alpha)y$$

$$\Rightarrow \int(\beta,\alpha)d\beta = \int \frac{|V(\alpha)|^{-1/2}}{(2\pi)^{n/2}}exp\{\frac{1}{2}(y - X\hat{\beta}(\alpha))^t V(\alpha)^{-1}(y - X\hat{\beta}(\alpha))\}\times\frac{(2\pi)^{n/2}}{|A(\alpha)^{-1}|^{-1/2}}$$

$$l_{(R(\alpha))} = \frac{1}{2}\{In|V(\alpha)| + (y - X\hat{\beta}(\alpha))^t V(\alpha)^{-1}(y - X\hat{\beta}(\alpha))\} - \frac{1}{2}In|A(\alpha)| + K$$

We now let

$$l_{(p(\alpha))} = \frac{1}{2}\{In|V(\alpha)| + (y - X\hat{\beta}(\alpha))^t V(\alpha)^{-1}(y - X\hat{\beta}(\alpha))\}$$

Such that

$$l_{(R(\alpha))} = l_{(p(\alpha))} - \frac{1}{2}In|A(\alpha)| + K \tag{3.7}$$

The restricted maximum likelihood of α is given by $\hat{\alpha}$ which maximizes the marginal log likelihood function in equation (3.7).

Estimation of β and υ

Since β is the set fixed effects fixed effects parameters and $V = ZGZ^T + R$ we make use of the GLS estimation procedure to find the Best Linear Unbiased Estimates (BLUE). "The GLS estimator properly takes into account the residual heteroscedasticity and correlation among residuals,"(Mehttalo, 2013). Assuming the variance component of the model are uncorrelated with each other, the least square equation for estimating β is given by:

$$\hat{\beta} = (X'X)^(-1)X'y$$

In the place of standard linear regression where $var(e) = \delta_e^2 I_n$ we have the variance of the mixed model given as var()=V. The GLS estimates of beta are given by:

We could also solve equation the array of matrices below to yield the same results. Here we recall that X is an $n \times p$ design matrix with respect to the fixed effects of the model. Z is also a design matrix for the random effects with dimension $n \times q$.

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ \upsilon \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix}$$

To show that the GLS estimate of the fixed effects parameters are unbiased,

$$E(\hat{\beta}) = E((X^T V^{(-1)} X)^{-1} X^T V^{-1} y)$$

$$= (X^T V^{-1} X)^{-1} X^T V^{-1} E(y)$$

$$= (X^T V^{-1} X)^{-1} X^T V^{-1} X\beta$$

$$= \beta$$

$$cov(\hat{\beta}) = cov((X^T V^{-1} X)^{-1} X^T V^{-1} y)$$

$$= ((X^T V^{-1} X)^{-1} X^T V^{-1}) cov(y) ((X^T V^{-1} X)^{-1} X^T V^{-1})'$$

$$= \delta^2 [(X^T V^{-1} X)^{-1} X^T V^{-1} V V^{-1} X (X^T V^{-1} X)^{-1}]$$

$$= \delta^2 [(X^T V^{-1} X)^- 1 X^T V^{-1} X (X^T V^{-1} X)^{-1}]$$

$$= \delta^2 [(X^T V^{-1} X)^{-1}]$$

Next, we estimate the Best Linear Unbiased Predictor of the random effects parameter by recalling that

$$y \sim N_n(X\beta, V); \quad \upsilon \sim N_{mq}(0, G); \quad \varepsilon \sim N_n(0, R); \quad and \quad cov(\upsilon, \varepsilon) = 0$$

$$cov(y, \upsilon) = cov(X\beta + Z\upsilon + \varepsilon, \upsilon)$$

$$= cov(X\beta, \upsilon) + Zvar(\upsilon, \upsilon) + cov(\varepsilon, \upsilon)$$

$$= ZG$$

Given these assumptions, we can establish a joint distribution for the two random variables such that:

$$\begin{pmatrix} y \\ \upsilon \end{pmatrix} \sim N_{n+mq} \left(\begin{pmatrix} X\beta \\ 0 \end{pmatrix}, \begin{pmatrix} V & ZG \\ GZ^T & G \end{pmatrix} \right)$$

The marginal distribution of υ is given as

$$\upsilon | y \sim N(\mu_{u|y}, \Sigma_{y|u}); \qquad \mu_{u|y} = \mu_\upsilon + \Sigma_{uy} \Sigma_y{}^{-1} (y - \mu_y); \qquad \Sigma_{y|u} = \Sigma_{uy} \Sigma_y^{-1} \Sigma_{yu}$$

The expectation of the conditional distribution of υ is computed as

$$E(\upsilon | y) = \mu_\upsilon + \Sigma_{uy} \Sigma_y^{-1} (y - \mu_y)$$

$$GZ^T V^{-1} (y - X\hat{\beta}) = \hat{\upsilon}$$

3.3.4 Hypothesis Testing

In linear mixed effects models hypotheses can be tested on the significance of the fitted model as well as on individual parameters. “Asymptotic results on the distribution of the maximum likelihood estimators and the restricted maximum likelihood estimators are used to derive confidence intervals and hypotheses tests for the model's parameters.”(Pinheiro and Bates, 2000) “The earliest estimation methods in the context of mixed-effects models were based on an ANOVA- approach. The method lead to unique, unbiased estimators for balanced datasets, where there is an equal number of observations in each group and no missing data.”(Mehttalo, 2013). Because of the unbalanced nature of the pupils score data, the Likelihood Ratio Test is used to examine the adequacy of the fitted model.

Testing Hypothesis on the Fitted Model

The fitted mode is assessed to establish its adequacy. Likelihood is the probability of seeing the data you collected given your model. The logic of the likelihood ratio test is to compare the likelihood of two models with each other. First, the model without the factor that youre interested in (the null model), then the model with the factor that youre interested in, (Winter, 2013). To obtain the null model, I make restrictions on and u of the fitted model such that the mean of the restricted model is zero and its variance is equal to ${}^2_e I$. The restricted model is then tested against the fitted model and the hypothesis is given by:

H_0: the null model is sufficient

H_a: the fitted model is significantly better than the null model.

Likelihood Ratio Test statistic used in making the decision on the adequacy of the fitted model is given by:

$$LRT = 2ln(L_2/L_1) = 2[In(L_2) - In(L_1)]$$

L_1 and L_2 are the likelihood functions of null and fitted models respectively. The likelihood function is given by:

$$L(y|\beta, \upsilon, \delta^2) = Pr(y|\beta, \upsilon, \delta^2)$$

$$(2\pi\delta^2)^{(-n/2)} exp\{\frac{1}{2\delta^2}\Sigma(y - X\beta)^2\}; \qquad \delta^2 = Z^T GZ + R$$

The likelihood ratio statistic has a chi square distribution with pq degrees of freedom i.e. $LRT \sim \chi^2(p-q)$. The likelihood ratio statistic is compared to the critical value of the chi square to make a decision either to reject or fail to reject the null hypothesis. The null hypothesis is rejected if the test statistic value is less than the critical value read from the chi square distribution.

Chapter 4

Data Analysis

4.1 Introduction

This chapter presents analysis of the results obtained using the model selected in chapter three. I have included some data manipulation processes in the chapter to show how the data that was used in the research project was created. Exploratory data analysis has been presented to give readers an insight into the data that has been used in the analysis. Various hypotheses have been tested and their significance discussed within the various sections.

4.2 Design of Study

Kenya has a total of forty seven counties. In each of the counties KCPE examinations are administered each year through schools that register candidates. The data used in this research are extract of 2013 KCPE results. I purposively selected Homa Bay County being my home county to demonstrate the use of mixed models in analysis of nested data. All the candidates who registered and sat for 2013 examination in the whole county have been considered. All the six districts in Homa Bay County have been considered in the research. The total number of schools selected was eight hundred and twenty six (826) being a sub set of all functional schools and registered examination centres in the county.

4.3 Data Management

I obtained the first dataset from the Kenya National Examination Council (KNEC) having the index of the pupil; their sex, their year of birth; their examination score during the 2013 KCPE; and the school from which they sat for their examination. This completed the pupil level variables. The second data set was obtained from the Ministry of education Science and Technology (MOEST) detailing the school name; the status of the school; the accommodation type; the category; the location of the school, a unique identifier for each school as well as the district within which the schools is found. These formed the second level of hierarchy data.

Harmonization of datasets

I carried out merging of the two datasets in Ms Excel making use of INDEX and MATCH functions. The Ms Excel is one of the most powerful data management tools that exist on the face of the earth. The flexibility with which it allows users to create or use existing formulas based on their needs is so incredible. Since the two data sets had school names, the first step was to merge the two using the schools.

This was however not as straight forward as it may seem as the characters used in the schools nomenclature were not necessarily the same so aliases were created. Manual identification of schools was employed at some point. In addition, some examination centres are not schools so a decision was made to leave them out of the analysis as there would be no corresponding schools characteristics for them. Once schools were matched, I supplied to each school in the KNEC list the corresponding unique school code from the MOEST data set for ease of further data management.

4.4 Model Diagnosis

For the mixed effects model to hold, there are couple of assumptions that the data under research need to meet. Like in the linear model, the linear mixed model applies similar assumptions of homoscedasticity, linearity, normality, independence.

Homoscedasticity

One graphical summary that should be examined routinely is a plot of the residuals versus the fitted responses from the model. This plot is used to assess the assumption of constant variance of the $\varepsilon_i j$, (Pinheiro and Bates, 2000). Figure 4.1 is a plot of the raw residuals plotted against the fitted pupils score showing decreasing variance with increase in the pupils KCPE scores. Figure 4.2 is a plot of the standardized pupil level residuals against the fitted values showing a trend similar to the raw scores. The residuals of the plot in Figure 4.2 are standardized by computing respective residual for each observation and dividing through by the standard deviation of the residual.

Figure 4.1 and Figure 4.2 indicate obvious violation of the assumption of homoscedasticity. However, since the estimation of β was done using the GLS procedure which took into account possible heteroscedasticity in the variance, this requirement can certainly be relaxed.

Normality

The random and residual components of the model all together are assumed to

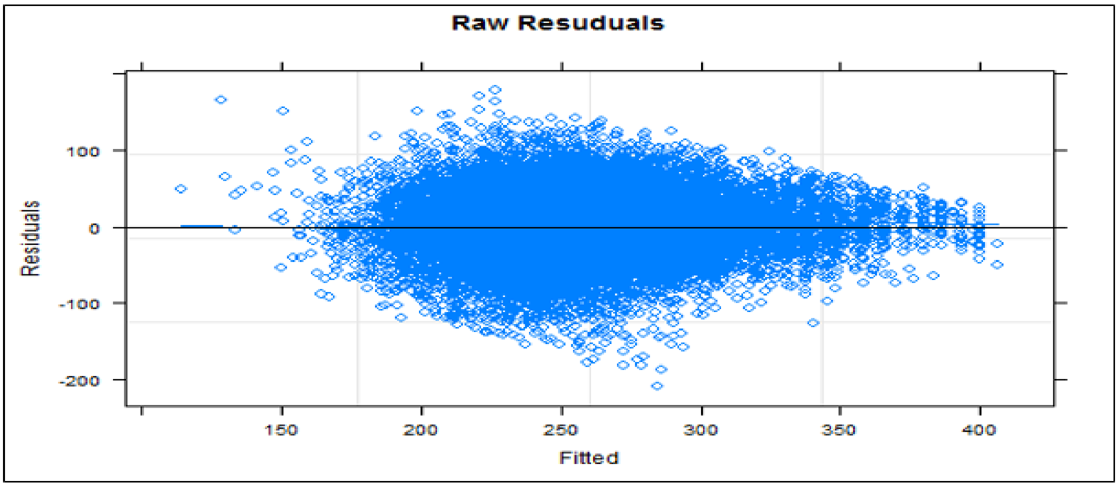


Figure 4.1: **Fitted Model vs Raw Residuals**

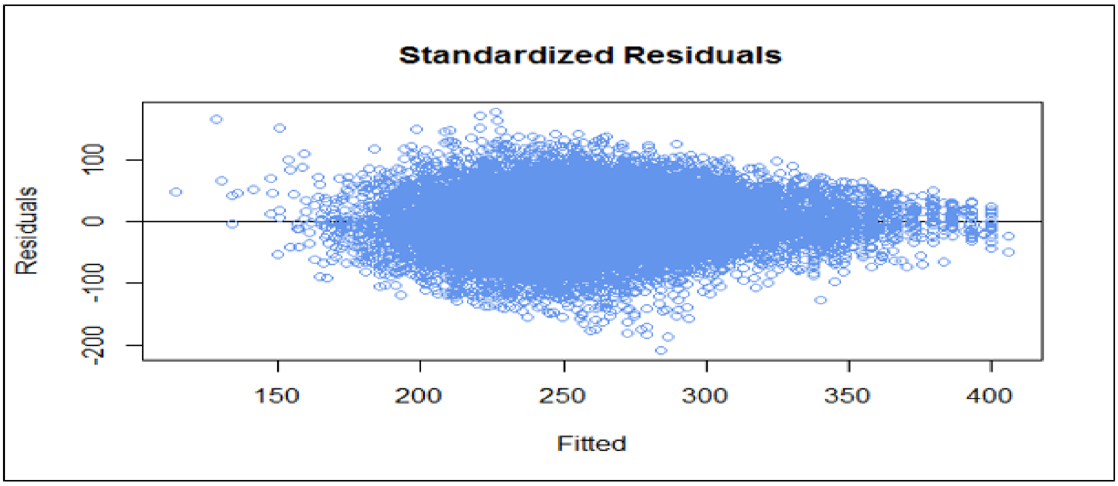


Figure 4.2: **Fitted Model vs Standardized Residuals**

be multivariate normally distributed. Figure 4.3 is a plot of the distribution of the residuals. The histogram to the left of Figure 4.3 suggests normal distribution which is confirmed by the Normal QQ-plot. The Assumption of normality in the distribution of the residuals is this sufficiently satisfied.

Linearity

The fitted model is tested for linearity as it is one of the assumptions made. Figure

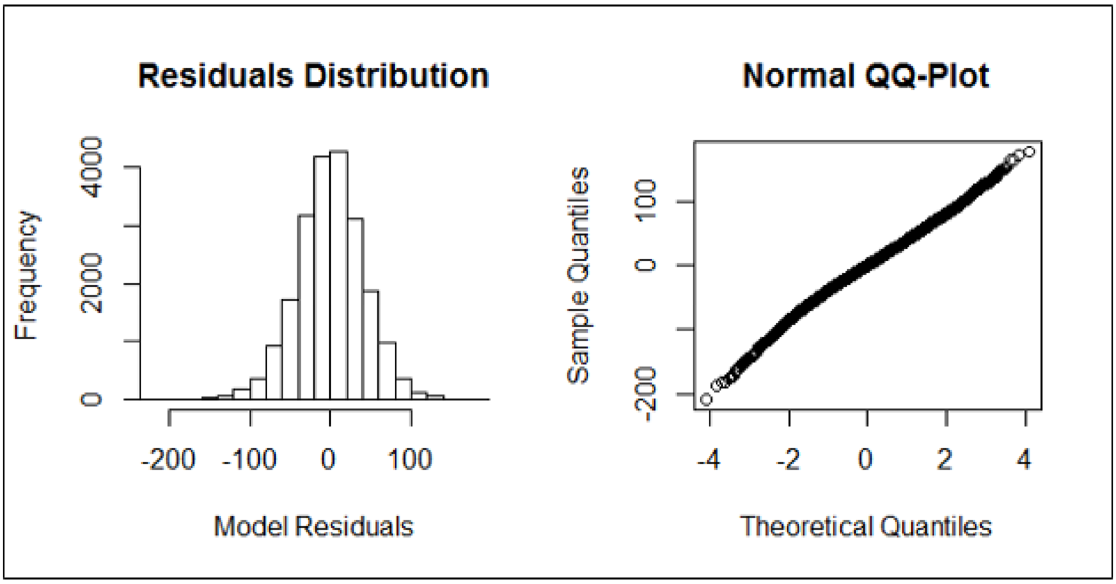


Figure 4.3: **Distribution of Model Residuals**

4.4 satisfies the assumption of normality as there is no indication of a polynomial trend. The KCPE scores is confirmed to be influenced by a linear combination of the fixed effects, the random effects and the error term.

Independence

One of the reasons why the mixed model is preferred to the linear model is its ability to address the non-independence of data. However, caution still needs to be taken to ensure that all variables picked for modeling are effective in the resultant model and cross check is done for misspecification.

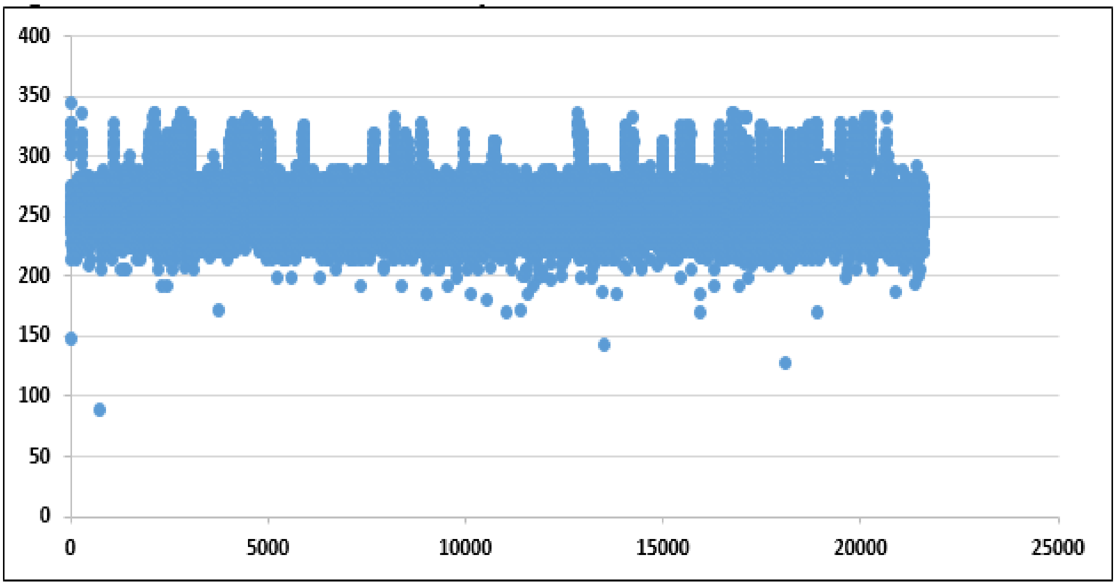


Figure 4.4: **Distribution of Model Residuals**

4.5 Exploratory Data Analysis

The study involved 826 schools drawn from the entire county with about 8% privately owned. The number of pupils considered for the study is 21,595. There are 9,784 girls accounting for 45.3% of the total sample. In terms of location, the number of pupils attending schools in urban areas accounted for 12.9% of the total sample. The Minimum age for the sample is 11 years with a maximum of 36. The Mean age of the learners is 14.99 years. The theoretical age for all grade eight pupils in Kenya is 13 years. A test of mean on the distribution of the age of the sampled pupils shows that there is a significant difference between their true mean and the expected mean.

During analysis, 210 observations are omitted from the study as the corresponding KCPEScores are missing. The Pupil with the least score KCPEScore in the sample managed 73 marks, 357 less than the pupil who scored the highest.

Table 4.1: **Summary of Data Variables**

Sex	Age	KCPE Score	Status	Location
Female: 9,784	Min. :11.00	Min. : 73.0	Public :19,876	Rural:18,812
Male :11,811	1st Qu.:14.00	1st Qu.:222.0	Private: 1,719	Urban: 2,783
	Median :15.00	Median :258.0		
	Mean :14.99	Mean :258.7		
	3rd Qu.:16.00	3rd Qu.:296.0		
	Max. :36.00	Max. :430.0		
		NA's :210		

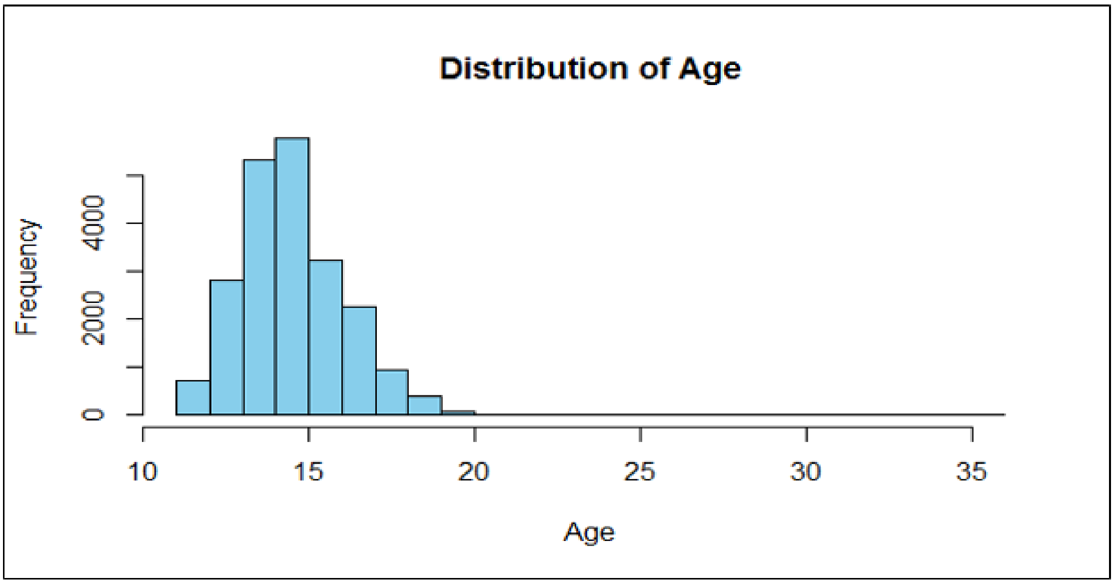


Figure 4.5: **Distribution of Pupils Ages**

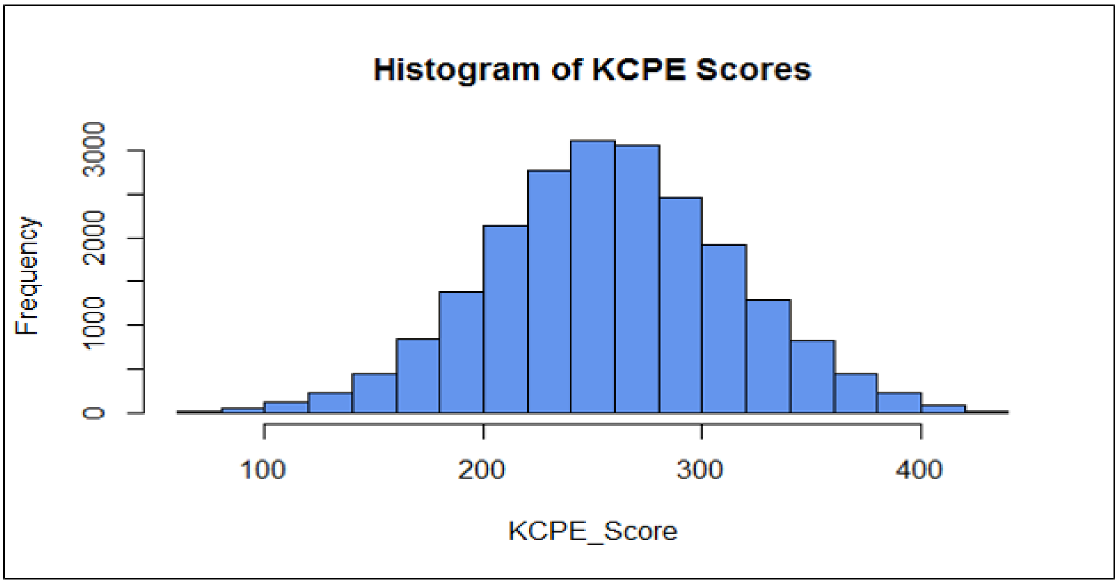


Figure 4.6: **Distribution of KCPE Scores**

4.6 Results

This section presents the results of the modelling. The study intended to develop an explanatory linear mixed effects model to sufficiently explain the varying phenomenon on the results of pupils'score in the national examinations. Specific objectives to the study included establishing the grouping effect of schools on the variations of pupil scores and to estimate the proportion of variance accounted for by schools. The model fitted satisfied the assumptions made except the assumption of homoscedasticity of residuals whose correction has been handled.

4.6.1 Model Adequacy

As noted in chapter three, the adequacy of a linear mixed effect model is established by a likelihood ratio test. The fitted model is compared to a null model whose parameters are all restricted. The R output for the likelihood ratio test is given in Table 4.We recall that the hypotheses tested for model adequacy here are:

H_0: *Null model is sufficient*

H_a: *Fitted model is significantly better than null model*

Table 4.2: **Likelihood Ratio Test Output**

Model	Df	LogLik	Df	Chisq	Pr(¿Chisq)
Null Model	3	-112,348.98	NA	NA	NA
Full Model	13	-111,196.87	10	2,304.21	0

From the results, the LRT test statistic is significant with a value of 2304.21. I reject the null hypothesis and conclude that the full model is better than the null model.

4.6.2 Model Results And Inference

Linear mixed model fit by REML ['lmerMod']

Formula: KCPEScore Age * Location + Age * Status + Sex * Location + Status * Sex + (Location—School)

REML criterion at convergence: 222,393.7

The output given by running the model is broken into sections. The first part of the output recalls what business went on in the model fitting. It shows that the model is of the Linear Mixed Model and was fitted using the Restricted Maximum Likelihood method. Since the data was run in R, the output further indicates that the fitting was made possible by the lmer function from the lme4 package. The formula (model) that has been fitted is printed out output of the model is a confirmation that the model was fitted using the Linear Mixed Effects Regression

(LMER/lmer) and the formula fitted is given alongside.

Randon Effects

The grouping effects of schools is quite visible as schools account for about 30% of the variations in the pupils scores. This is given by the inter-class-correlation. In the case of **Fixed Effects**

Table 4.3: **Random Effects Estimates**

Groups	Name	Variance	Std.Dev.	Correlation
School	(Intercept)	778	27.89	
	Location:Urban	590.1	24.29	-0.49
Residual		1,753.8	41.88	

Table 4.4 gives the fixed effects estimates of the fitted model based on the R

Table 4.4: **Fixed Effects Estimates**

Effect	Estimate	Standard Error	t value
(Intercept)	352.4157	3.1825	110.74
Age	-7.2955	0.199	-36.65
LocationUrban	36.0963	9.4435	3.82
StatusPrivate	41.2511	11.593	3.56
SexMale	16.7453	0.6549	25.57
Age:LocationUrban	-1.7861	0.5926	-3.01
Age:StatusPrivate	0.5265	0.7658	0.69
LocationUrban:SexMale	-4.2269	1.7589	-2.4
StatusPrivate:SexMale	-3.2869	2.2117	-1.49

output. The output does not display the p-values corresponding to the estimates. However since the sample is large, the t-distribution follows a normal distribution and the t-values cab be directly compared to the 1.96 from the normal distribution at (α=0.05) confidence. With this in mind we note that interaction between school status and pupils age as well as the interaction between school status and pupils sex are not significant as their corresponding t-values are lower than the critical value of 1.96.

i. The grand mean is 352.42 implying that if all the explanatory variables are controlled, a pupil would score 352.42.

ii. A unit increment in the age of a pupil in urban location causes a decrease in the KCPEScore by 1.79.

iii. A male pupil in urban location scores 3.2 marks less

iv. A unit increment in age in private school increases the KCPE Score by 0.53

Since the interactions are significant the main effects of the model are not interpreted much as R provides them as outputs.

Correlation of Fixed Effects

The final section of the output provides a correlation matrix that highlights the correlation between fixed effects variables of the fitted model. The coefficients in the matrix are quite low in all cases pointing to independence of the explanatory variables (absence of multicollinearity).

Table 4.5: **Correlation of Fixed Effects**

	(Intr)	Age	LctnUr	SttsPr	SexMal	Ag:LcU	Ag:StP	LcU:SM
Age	-0.932							
LocatinUrbn	-0.295	0.282						
StatusPrivt	-0.223	0.211	-0.096					
SexMale	-0.024	-0.096	0.004	0				
Ag:LctnUrbn	0.282	-0.305	-0.921	0.069	0.032			
Ag:SttsPrvt	0.2	-0.215	0.065	-0.946	0.024	-0.067		
LctnUrbn:SM	0.005	0.034	-0.027	0.02	-0.327	-0.076	-0.008	
SttsPrvt:SM	0.001	0.026	0.02	-0.015	-0.226	-0.009	-0.089	-0.119

4.6.3 Plots of KCPE Performance by Selected Variables

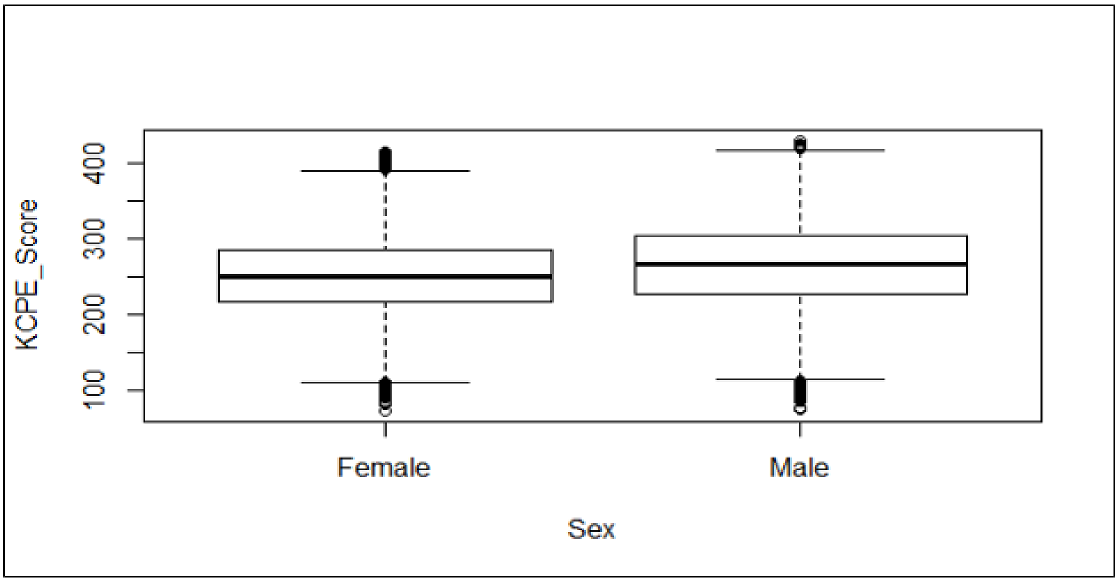


Figure 4.7: **Performance by Sex**

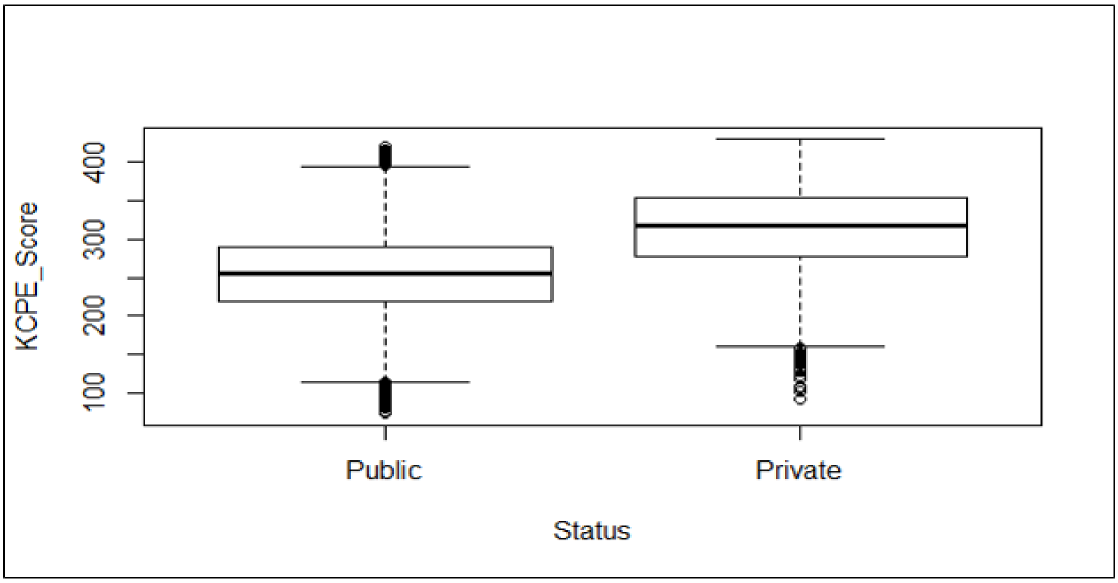


Figure 4.8: **Performance by Status**

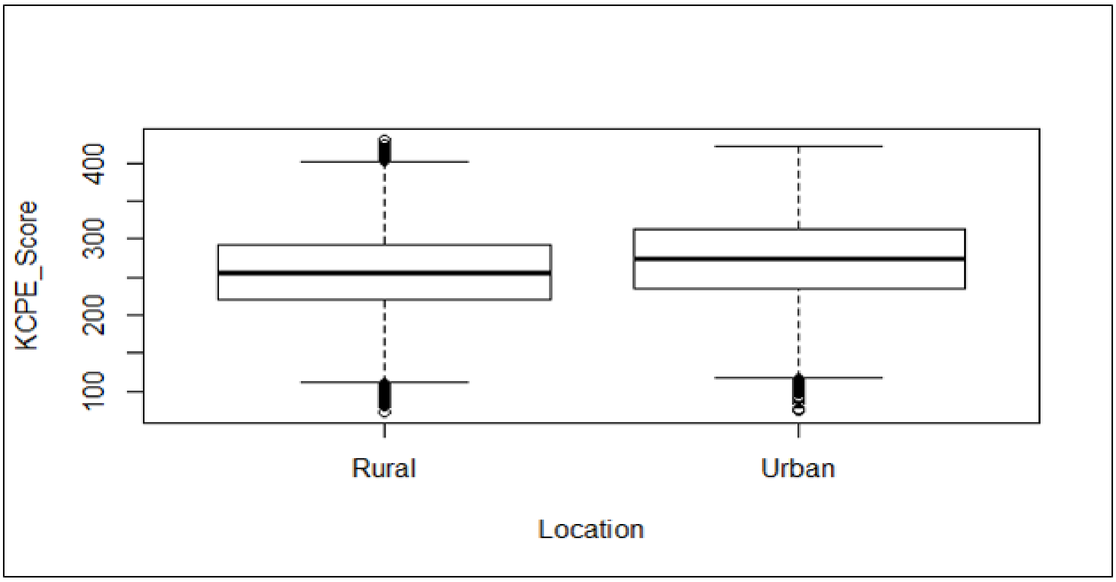


Figure 4.9: **Performance by Location**

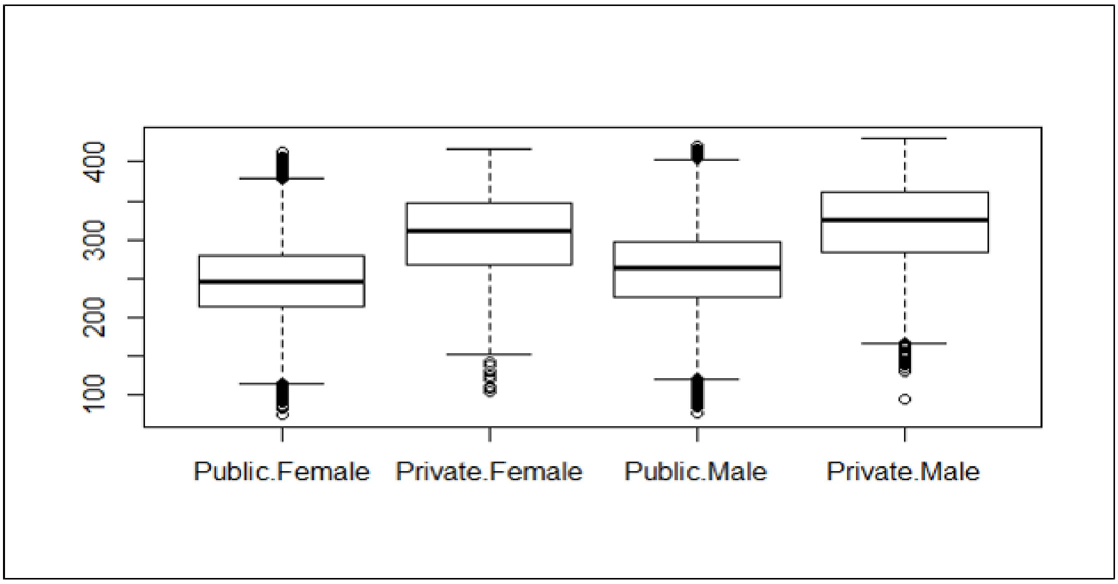


Figure 4.10: **Performance by Interaction of Sex and Status**

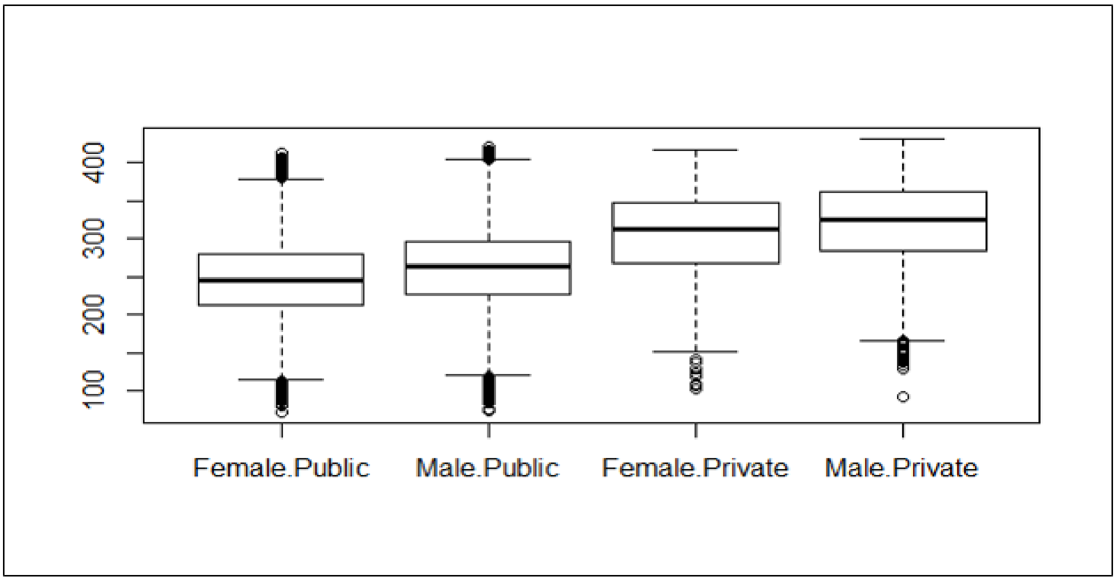


Figure 4.11: **Performance by Interaction of Status and Sex**

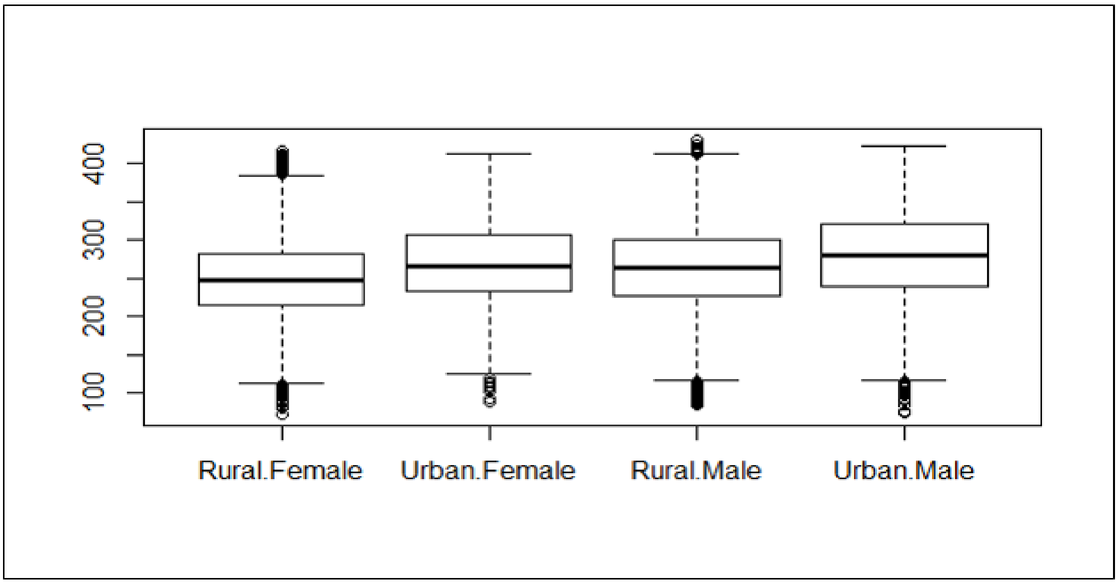


Figure 4.12: **Performance by Interaction of Sex and Location**

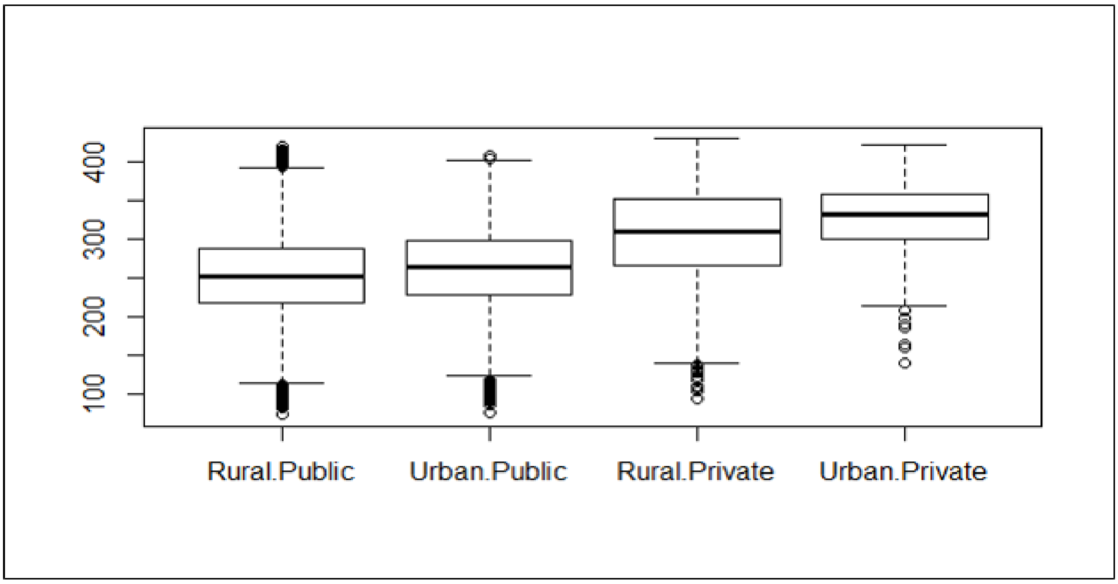


Figure 4.13: **Performance by Interaction of Status and Location**

4.7 Discussion

It is quite evident from the results of the model that the variations in the examinations are not entirely due to individual difference in capability. There are other factors that influence the variation which are the random systemic occurrences. This study has established that 30% of the total variation is due to the schools group effect. With a correlation of -0.49, the inclusion of the random effects in the model is justified and upheld. This study thus makes comparison with the findings of (Gustafsson, 2007). Using hierarchical modeling, he established that the structure of historically disadvantaged schools was a major reason accounting for the variation between schools.

In the Kenyan case this may just be the reason. The education system over the years has treated schools in a similar manner. There is a systemic marginalization of some schools while others are made to thrive. Schools that have had their management close to perceived political power have been rewarded with better infrastructure; better financial resources; and other human inputs that facilitate learning. In the South African case, (Gustaffson, 2007) implores the Government to increase funding to schools that were traditionally neglected by the administrations that presided over the country during and after the apartheid ages. In a similar fashion, Kenya has a chance of correcting the historical bias and work towards improving the entire education system. Schools that are already performing well should be sustained to continue performing well. Schools with systemic weaknesses should be supported more to ensure that they catch up with the schools that perform well. This will see the system achieve joint common gains.

Chapter 5

Conclusions and Recommendations

5.1 Introduction

This chapter closes the study with a summary of the findings of the research as well as recommendations that will further the objectives of this study.

5.2 Conclusions

The study has established in line with the aim that schools account for significant proportion of the variations observed in pupils scores. Schools have been established to account for 30% of the variation. The pupils account for about 45% of the variation. This shows that over half of the disparities observed in assessments can be addressed through systemic adjustments. Gender disparity has come out very strongly. Rural schools perform relatively lower than urban schools. Private schools perform relatively better than public schools.

5.3 Recommendations

The model that was used in the study was a two level model. With the findings that have been made, there is a strong possibility that higher level hierarchies could still account for more proportions of the variations. With the County Governments having been operationalized, this could be an impetus to call them to action in improving education in counties that may be found wanting. I recommend that more study be carried to establish whether there are other factors explaining the variations. In addition to the higher level hierarchies, there is need to look at the cause of within schools variation beyond the age and the sex of pupils. Household characteristics may be big influence of the scores and may need to find themselves in the list of variables.

My final recommendation is the adoption of this procedure in the analysis of assessments such that beyond the subjects/pupils, the effect of their environment may be taken into just consideration. This way it there will be a rational attempt to apply equalizers when there are known causes of inequality.

References

1. ACER. (2015). Southern and Eastern Africa Consortium for Monitoring Educational Quality, 18. Retrieved from http://research.acer.edu.au/cgi/

2. Braun, H., and Kanjee, A. (2006). *I*mproving education through assessment , innovation , and evaluation. Retrieved from https://www.amacad.org/publications/braun.pdf

3. Davidian, M. (n.d.). Linear Mixed Effects Models for Multivariate Normal Data (pp. 363422).

4. Foshay, A. W. (1991). The curriculum matrix transcendence and mathematics, 6(4), 277293.

5. Goldstein, H. (1999). *M*ultilevel Statistical Models (Vol. 3).

6. Gustafsson, M. (2007). *U*sing the hierarchical linear model to understand school production in South Africa (No. 01/07). Retrieved from http://www.ekon.sun.ac.za/wp 01-2007.pdf

7. Kachman, S. D. (2000). An Introduction to Generalized Linear Mixed Models (pp. 5973).

8. Kastberg, D., Roey, S., Lemanski, N., Chan, J. Y., and Murray, G. (2014). *T*echnical Report and User s Guide for the Program for International Student Assessment (PISA).

9. King, M. L. J. (1947). *T*he Purpose Of Education.

10. Kirchner, J. (2001). Simple linear regression.

11. Kyle, R. J. (2002). Nested ANOVA vs. crossed ANOVA: when and how to tse which, (2).

12. Mehttalo, L. (2013). Linear Mixed-Effects Models with Examples in R.

13. OECD. (2014). *O*ECD (2014), PISA 2012 Results: What Students Know and Can Do Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014), PISA, OECD Publishing (Vol. I). Retrieved from http://dx.doi.org/10.1787/9789264201118-en ISBN

14. Ondima, C. M., Nyamasege, D., Mogwambo, V. A. and Ochoti, G. N. (2013). Regression analysis of entry scores (KCPE) and final performance (KCSE) in Kenya: a case of Nyamira District, Nyamira County, Kenya, 3(4), 718.

15. Onyango, N. O. (2009). On the linear mixed effects regression (lmer) R function for nested animal breeding data. *C*s-Bigs, 4(1), 4458. Retrieved from http://www.bentley.edu/csbigs/documents/onyango.pdf

16. Pinheiro, J. C. and Bates, D. M. (2000). *M*ixed-Effects Models in S and S-Plus.

17. Piper, B. (2010). *K*enya early grade reading assessment report. Retrieved from https://www.eddataglobal.org/reading/index.cfm/EGRA Kenya 072910 Final.pdf

18. Rencher, A. C. and Schaalje, B. G. (2008). *L*inear Models in Statistics (2nd ed.).

19. Taylor, J. (n.d.). Introduction to regression and analysis of variance: fixed vs random effects, 119.

20. UNICEF. (2000). *D*efining quality in education (No. UNICEF/PD/ED/00/02).

21. Uwezo Kenya. (2013). *A*nnual learning assessment report 2012.

22. Vlaardingerbroek, B., Taylor, N. and Heyneman, S. (2008). External Examinations Systems.

23. Wasanga, P. M., Ogle, M. A.and Wambua, R. M. (2012).

24. The SACMEQ III project in Kenya: a study of the conditions of schooling and the quality of education. SACMEQ educational policy research series. Retrieved from http://www.sacmeq.org/sites/default/files/sacmeq/reports/sacmeq-iii/national-reports/kenyanationalreport.pdf

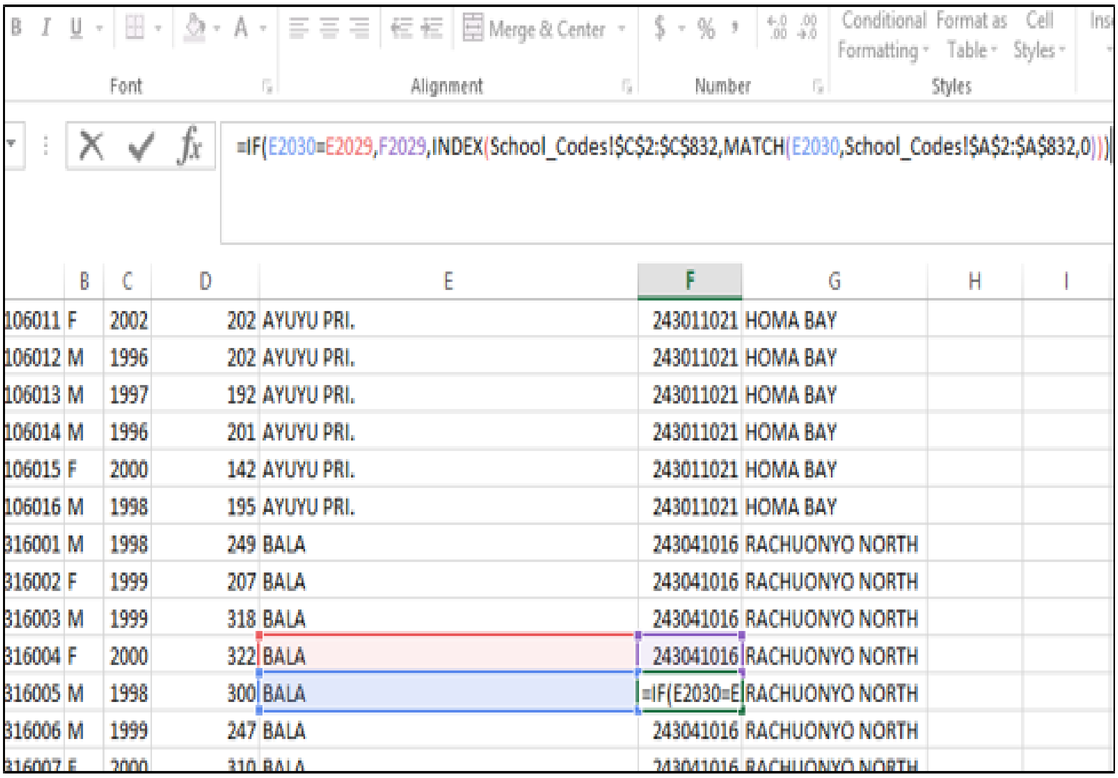
25. Winter, B. (2006). Distinguishing between random and fixed variables, effects, and coefficients, 13.

26. Winter, B. (2013). Linear models and linear mixed effects models in R, 122.

27. Woltman, H., Feldstain, A., Mackay, J. C. and Rocchi, M. (2012). An introduction to hierarchical linear modeling. *T*utorials in Quantitative Methods for Psychology, 8(1), 5269. Retrieved from http://www.tqmp.org/RegularArticles/vol08-1/p052/p052.pdf

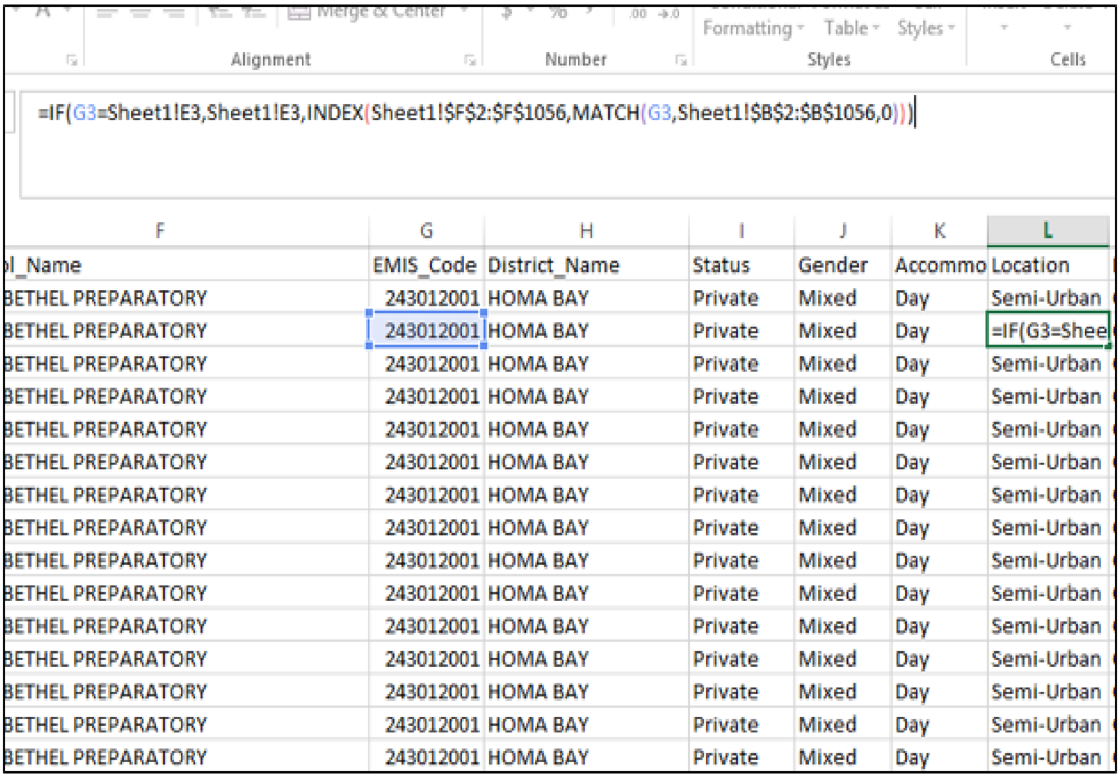
Appendices

Data Management in MS Excel



	B	C	D	E	F	G	H	I
106011	F	2002	202	AYUYU PRI.	243011021	HOMA BAY		
106012	M	1996	202	AYUYU PRI.	243011021	HOMA BAY		
106013	M	1997	192	AYUYU PRI.	243011021	HOMA BAY		
106014	M	1996	201	AYUYU PRI.	243011021	HOMA BAY		
106015	F	2000	142	AYUYU PRI.	243011021	HOMA BAY		
106016	M	1998	195	AYUYU PRI.	243011021	HOMA BAY		
316001	M	1998	249	BALA	243041016	RACHUONYO NORTH		
316002	F	1999	207	BALA	243041016	RACHUONYO NORTH		
316003	M	1999	318	BALA	243041016	RACHUONYO NORTH		
316004	F	2000	322	BALA	243041016	RACHUONYO NORTH		
316005	M	1998	300	BALA	=IF(E2030=E	RACHUONYO NORTH		
316006	M	1999	247	BALA	243041016	RACHUONYO NORTH		
316007	F	2000	310	BALA	243041016	RACHUONYO NORTH		

Figure 5.1: **Assigning Pupils A School Unique Code**


=IF(G3=Sheet1!E3,Sheet1!E3,INDEX(Sheet1!$F$2:$F$1056,MATCH(G3,Sheet1!$B$2:$B$1056,0)))

F	G	H	I	J	K	L
l_Name	EMIS_Code	District_Name	Status	Gender	Accommo	Location
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	=IF(G3=Shee
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban
BETHEL PREPARATORY	243012001	HOMA BAY	Private	Mixed	Day	Semi-Urban

Figure 5.2: **Assigning School Attributes to Pupils**

Appendix 2: R-Codes

Loading Data

kcpe=read.csv("Pupils Scores.csv")

kcpe.sub$Location = factor(kcpe.sub$Location,levels=c(1,2),labels=c("Rural","Urban"))

kcpe.sub$Status = factor(kcpe.sub$Status,levels=c(1,2),labels=c("Public","Private"))

kcpe.sub$Sex = factor(kcpe.sub$Sex,levels=c("1",'2'),labels=c("Female","Male"))

kcpe.sub$School = factor(kcpe.sub$School)

attach(kcpe.sub)

library(plyr);library(doBy);library(nlme);library(R2wd)

Models Fitting

model1=lmer(KCPEScore Age*Location+Age*Status+Sex*Location+Status*Sex+(Location—Sch

model2=lmer(KCPEScore Age*Status+Sex*Location+Status*Sex+(Location—School))

model3=lmer(KCPEScore Age+Sex*Location+Status*Sex+(Location—School))

model4=lmer(KCPEScore Age+Status*Sex+(Location—School))

model5=lmer(KCPEScore Age+Sex+Location+Status+(Location—School))

model6=lmer(KCPEScore Age+Sex+Location+Status+(1—School))

Model selection

anova(model1,model2,model3,model4,model5,model6)

Posting Results Direct to MS Word using R2wd package

wdGet()

wdTable(anova(model1,model2,model3,model4,model5,model6))

Model Diagnosis **Homoscedasticity**

par(mfrow=c(1,2))

plot(model1,type=c("p","smooth"),main="Raw

Resuduals",xlab="Fitted",ylab="Residuals")

plot(model1,sqrt(abs(resid(.))) fitted(.), type=c("p","smooth")) plot(fitted(model1),residuals(model1),main="Standardized Residuals" ,xlab="Fitted",ylab="Resid

Normaity test

par(mfrow=c(1,2))

hist(residuals(model1),main="Residuals Distribution",xlab="Model Residuals")

qqnorm(residuals(model1),main="Normal QQ-Plot",abline(a=0,b=0))

par(mfrow=c(1,1))

Model adequacy: creation of a null model

null=lmer(KCPEScore 1+(1—School))

summary(null)

anova(null,model1)

Likelihood Ratio Test

library(lmtest)

lrtest(null,model1)

Exploratory data analysis

Summary=ddply(kcpe.sub,c("Location","Status"),summarize, Pupils=length(School),

Boys=round(mean(KCPEScore[Sex=="Male"],na.rm=T),digits=2), Girls=round(mean(KCPEScor

wdTable(Summary) ***Test on means of Age***

t.test(Age,mu=13)

hist(Age, breaks=25,main="Distribution of Age",col="sky blue");

hist(KCPEScore, breaks=20, col="cornflowerblue", xlab="KCPEScore", main="Histogram of KCPE Scores")

Boxplots

plot(KCPEScore Sex)

plot(KCPEScore Location)

plot(KCPEScore Status)

boxplot(KCPEScore Status+Sex)

boxplot(KCPEScore Sex+Status)

boxplot(KCPEScore Location+Sex)

boxplot(KCPEScore Location+Status)

boxplot(KCPEScore Status+Location)

summary of the fitted model Random and Fixed Effects Estimates

summary(model1)

wdTable(model1)

print(vc = VarCorr(model1), comp = c("Variance","Std.Dev."))

fixef(model1)

Schools Random Intercepts and Slopes

ranef(model1)

coef(model1)

Predictive power of the model

iqrvec = sapply(simulate(model1,3000),IQR)

obsval=IQR(KCPEScore,na.rm=T)

post.pred.p=mean(obsval>=c(obsval,iqrvec))

post.pred.p