UNIVERSITY OF GHANA

COLLEGE OF BASIC AND APPLIED SCIENCES
SCHOOL OF PHYSICAL AND MATHEMATICAL SCIENCES

LONGITUDINAL ANALYSIS OF CEREAL YIELDS IN GHANA

BY

ALFRED KWABENA AMOAH

(10507283)

THIS THESIS IS SUBMITTED TO THE UNIVERSITY OF GHANA, LEGON IN
PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF
MASTER OF PHILOSOPHY IN STATISTICS DEGREE

JULY, 2016
DECLARATION

I hereby declare that this thesis is the result of my own research work and that no part of it has been presented for another degree in this university or elsewhere.

SIGNATURE……………………………                                             DATE:......................

ALFRED KWABENA AMOAH
(10507283)

Supervisors’ Declaration

We hereby certify that this thesis was prepared from the candidate’s own research work and supervised in accordance with the guidelines on the supervision of thesis laid down by the University of Ghana.

SIGNATURE………………………………..                                       DATE:…………………

DR. SAMUEL IDDI
(PRINCIPAL SUPERVISOR)

SIGNATURE………………………………..                                     DATE:…………………

DR. ISAAC BAIDOO
(CO-SUPERVISOR)
ABSTRACT

The aim of this study was to investigate cereal crop yields in Ghana. Specifically, this research determines whether there is a significant difference in the yields across the ten regions in Ghana and also finds out the evolution of crop yields among the regions. Data on two major cereal crops (Maize and Rice) produced and consumed in Ghana was attained from Ministry of food and Agriculture (MOFA). Multivariate Analysis of Variance (MANOVA) and Linear Mixed model (LMM) were employed for the study. Diagnostic plots for the fitted Linear Mixed Model and MANOVA revealed a valid model for the analysis.

The study revealed that significant differences exist in the yields of the two major cereal crops in all the regions in Ghana. The study identified that significant differences occurred in the average yields of most regions’ yields with other regions for both maize and rice.

Further analysis by LMM indicated that the yields of maize and rice varied between and within the regions of Ghana. It also indicated that there is decelerating trend in maize yields and gradual increasing trend in rice yields across all the regions in Ghana. Based on these findings, we recommend that intensive support in the form of credit facilities and farm inputs must be given to farmers who engage in cereal crops production in all the regions in Ghana to help reduce this variability in the two major cereal crop yields. Also maize production especially must be encourage in all the regions to help avenge the declining trend in the yields of maize as it is considered the most consumed cereal crop in Ghana.

We further recommend the use of Joint Models to simultaneously study the trend of crops and also study factors such as rainfall and climate data which may influence cereal crop production in Ghana.
DEDICATION

I dedicate this special study to my Dear wife Ernestina Amoah and my son Harvey Adjei-Tachie Amoah.
ACKNOWLEDGEMENT

I would like to express my candid and sincere thanks to Dr. Samuel Iddi and Dr. Isaac Baidoo, Lectures at Statistics Department of University of Ghana whose supervision, helpful comments, suggestions and incessant advice spurred me to carry out the work successfully.

I wish to place on record my heartfelt thanks to all the lecturers of Department of Statistics whose tuition equipped me with prerequisite knowledge required for this study as well as workers at SRID of Ministry of Food and Agriculture for helping me with the data.

I cannot also forget the support I received from Madam Felicitas Agana of Bank of Ghana, Sunyani, Solomon Owusu-Bempah of Ghana Statistical Service, Head Office and Sarah Kwakye my mate at UG for her encouragement as well as all my friends.

Finally, I sincerely acknowledge the support given to me by my dear wife, Ernestina Amoah throughout this study. Dear, may God richly bless you for a wonderful encouragement and support given to me.
**TABLE OF CONTENTS**

DECLARATION ................................................................................................................... i  
ABSTRACT ......................................................................................................................... ii  
DEDICATION ..................................................................................................................... iii  
ACKNOWLEDGEMENT ....................................................................................................... iv  
TABLE OF CONTENTS ...................................................................................................... v  
LIST OF TABLES ............................................................................................................. viii  
LIST OF FIGURES ............................................................................................................ ix  
LIST OF ABBREVIATIONS ............................................................................................... x  
CHAPTER ONE ................................................................................................................... 1  
INTRODUCTION ................................................................................................................ 1  
  1.0. Background of the study ....................................................................................... 1  
  1.1. Motivation of the research .................................................................................. 3  
  1.2. Problem Statement ............................................................................................. 4  
  1.3. Objective ............................................................................................................. 5  
  1.4. Research Questions ............................................................................................ 5  
  1.5. Significance of the study .................................................................................... 5  
  1.6. Scope and Methodology ...................................................................................... 6  
CHAPTER TWO .................................................................................................................. 8  
LITERATURE REVIEW ...................................................................................................... 8  
  2.0. Introduction ......................................................................................................... 8  
  2.1. Theoretical Framework ....................................................................................... 8  
  2.2. Conceptual Framework ...................................................................................... 11  
    2.2.1. A mixed effects model of crop yield ............................................................. 11  
    2.2.2. Modeling for crop yield in Ghana ................................................................. 16  
    2.2.3. The uniqueness of the study ......................................................................... 18  
CHAPTER THREE ............................................................................................................. 20  
METHODOLOGY .............................................................................................................. 20  
  3.0. Introduction ......................................................................................................... 20
4.4.4. Fixed effects parameter estimates ................................................................. 55
4.5. Discussion ........................................................................................................... 60
CHAPTER FIVE .......................................................................................................... 62
CONCLUSION AND RECOMMENDATION ......................................................... 62
5.0. Introduction ....................................................................................................... 62
5.1. Conclusion ........................................................................................................ 62
5.2. Recommendation ............................................................................................ 63
REFERENCE .......................................................................................................... 64
APPENDIX ................................................................................................................ A
Appendix A: Test of Assumptions ............................................................................. A
Appendix B: Test on significant differences of maize and rice yields ..................... D
Appendix C: Approximate 95% confidence intervals .............................................. F
Appendix D: Marginal Variance Covariance Matrix .................................................. G
Appendix E: Pairwise Comparisons of difference in Mean Yields ......................... H
Appendix F: R Codes used for the study ................................................................. K

University of Ghana http://ugspace.ug.edu.gh
LIST OF TABLES

Table 4.1: Descriptive statistics of the two major cereal crops……………………………43

Table 4.2a: MANOVA Test for Differences in crop yields across the ten regions......... 47

Table 4.2b: Test of Significant difference in crop yields........................................47

Table 4.3: Pairwise Comparisons of differences in mean yields.............................49

Table 4.4: Selection of best model...............................................................51

Table 4.5: Random effects parameter estimates................................................54

Table 4.6a: Fixed effects parameter estimates for maize yields.............................58

Table 4.6b: Fixed effects parameter estimates for rice yields...............................59
LIST OF FIGURES

Figure 4.1: Line graph showing average yields by year……………………………………….43

Figure 4.2: Boxplot of preliminary analysis .....................................................................44

Figure 4.3: Mean profile of maize yields per region.......................................................44

Figure 4.4: Mean profile of rice yields per region.............................................................45

Figure 4.5: Diagnostic plots for fitted Linear Mixed model..............................................53

Figure 4.6: Distribution of random intercepts (District) for maize and rice yields..........55
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>AKAIKE INFORMATION CRITERIA</td>
</tr>
<tr>
<td>AMMI</td>
<td>ADDICTIVE MAIN EFFECT AND MULTIPLICATIVE INTERACTION</td>
</tr>
<tr>
<td>ANOVA</td>
<td>ANALYSIS OF VARIANCE</td>
</tr>
<tr>
<td>BIC</td>
<td>BAYES INFORMATION CRITERIA</td>
</tr>
<tr>
<td>FAO</td>
<td>FOOD AND AGRICULTURAL ORGANIZATION</td>
</tr>
<tr>
<td>GDP</td>
<td>GROSS DOMESTIC PRODUCT</td>
</tr>
<tr>
<td>GSS</td>
<td>GHANA STATISTICAL SERVICE</td>
</tr>
<tr>
<td>KNUST</td>
<td>KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY</td>
</tr>
<tr>
<td>LMM</td>
<td>LINEAR MIXED MODEL</td>
</tr>
<tr>
<td>MANOVA</td>
<td>MULTIVARIATE ANALYSIS OF VARIANCE</td>
</tr>
<tr>
<td>MOFA</td>
<td>MINISTRY OF FOOD AND AGRICULTURE</td>
</tr>
<tr>
<td>NRDS</td>
<td>NATIONAL RICE DEVELOPMENT STRATEGY</td>
</tr>
<tr>
<td>OECD</td>
<td>ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT</td>
</tr>
<tr>
<td>SDG</td>
<td>SUSTAINABLE DEVELOPMENT GOALS</td>
</tr>
<tr>
<td>SRID</td>
<td>STATISTICS, RESEARCH AND INFORMATION DEPARTMENT</td>
</tr>
<tr>
<td>UK</td>
<td>UNITED KINGDOM</td>
</tr>
<tr>
<td>UN</td>
<td>UNITED NATION</td>
</tr>
<tr>
<td>USA</td>
<td>UNITED STATE OF AMERICA</td>
</tr>
<tr>
<td>US</td>
<td>UNITED STATE</td>
</tr>
</tbody>
</table>
CHAPTER ONE

INTRODUCTION

1.0. Background of the study

In the last two decades, the concern of production and the role of guaranteeing food security seem to have moved away from global focus. There has been a decline in investment in research, infrastructure and advancements in technology have decelerated (Rosegrant & Cline, 2003). Agricultural development has been limited and about to face its biggest challenges. “The past 35 years have seen food reserves been at its lowest because of the demand that faced food production at a time when crop yields are stagnant” (OECD & FAO, 2009; Premanandh, 2011; Rosegrant & Cline, 2003).

In 2008 the FAO, World Bank and G8 all called for a transformed investment in agriculture directed at improving and increasing agricultural productivity and production (OECD & FAO, 2009). The World Bank report (2009) highlights the global accessibility of agricultural land and the possible ways for improving production, particularly in areas with low agricultural development, such as Africa. Agreement appears to be establishing that “underutilized” land in developing countries can be the key to continuing global food safety, whether this is through the introduction of new farmland or the closing of yield gaps (FAO, 2009; Foley, 2011; Kearney, 2010; Tran, 2011; World, 2009). Unlike US or Europe where Agricultural prospective is very closed to maximized and soil quantity is degrading, advancing in the agricultural sectors of developing countries will improve productivity to a very large degree (Beddington, 2010). This investment in policies, technology, science and infrastructure is essential to access the potential productivity from the land and resources of
Africa (Bank, 2011; Rosegrant & Cline, 2003). The sense is that investment in developing nations, which are currently only accomplishing a small amount of their agricultural prospective, could contribute to future global food safety by increasing crop yields. This was reemphasized by UN General Assembly on 11th August, 2015 on the need to improve agriculture productivity by ensuring the attainment of Sustainable Development Goal 2 (SDGs 2) on hunger, achieved food security and improved nutrition and promotion of sustainable agriculture (Griggs et al., 2013).

In Ghana, we produce food crops such as maize, rice, plantain, cassava, yam and other vegetables are grown both on small scale and commercial level. “Annual production over the years continues to decline in relative term in spite of several programme interventions by the Municipal and District offices of the Ministry of Food and Agriculture” (Agyare et al., 2014). “This is due to increasing cost of farm inputs and the low soil fertility. Crop production is largely rain-fed and traditional technology of production continues to dominate the sector with peasant farmers using simple tools such as hoes and cutlasses. The average land holding per farmer is relatively low and is about 0.5 hectares” (Agyare et al., 2014).

Maize is one of the major cereals produced and consumed in Ghana. Majority of maize is produced by peasant farmers who use local farm inputs under rain fed conditions. Under this method of production, yields of maize have not been encouraging as it is below the attainable levels of 5.0-5.5 metric tons per hectare. The estimates of annual maize yields according to MOFA from 2010 to 2015 are between deficits of 84,000 to 145,000 metric tons over the last four years. This signifies a deficit in local production of between 9 and 15 percent of total consumptions in these years with only 2.6 annual growths. Rice follow closely as the second most important cereal crop consumed by most Ghanaians. It has average annual production yields of 1.0-2.4 metric tons per hectare. “Also between 2010 and 2015, rice demand is
projected to have grown at a compound annual growth of 11.8 percent from 939,920 metric tons to 1644,221 metric tons” (SRID, 2007).

Nevertheless, the country is not able to feed itself in either of its two most important cereal crops, as Ghana has experienced average downfalls in local maize supplies of 12 percent and domestic rice supplies of 69 percent in recent years (Agyare et al., 2014).

The yields of these two crops have therefore; become a concern to most agriculturists in Ghana with intensive demand on Ministry of Food and Agriculture (MOFA) to find a way of containing this problem. Over the years, Statistics, Research and Information Department (SRID) of MOFA have been analyzing the yields of all major crops in the country to find out the percentage changes (increases or decrease) that occur in these crops at each crop/farming season. In this thesis, we used multivariate analysis of variance (MANOVA) and linear mixed model (LMM) to analyze the yields of the two most important cereal crops (Maize and Rice) to investigate significant differences and trend over ten years among them across all the ten (10) regions in Ghana.

1.1. Motivation of the research

This motivation of the research is indicated by the fact that Agriculture contributes meaningfully to Ghana’s GDP and employs majority of Ghanaians especially the rural folks who form the majority of Ghana’s population. Modeling the differences in major food crops will thus help farmers to know which of the crops to increase productivity across the various districts of the region. Again, it is motivated by the need to redirect stake holders’ attention to the achievement of Sustainable Development Goal 2 (SDGs 2) on hunger, achieved food security and improved nutrition and promote sustainable agriculture (Griggs et al., 2013).
1.2. **Problem Statement**

The growth of population, development, and stationary agricultural production are backing to an incipient structural scarcity of food crops in the West Africa sub-region. “Finding ways of effectively coping with this emerging food deficit is critical for fostering economic growth, reducing poverty, and enhancing food/nutrition security for the people of West Africa. Addressing this challenge requires placing agriculture and the associated processes of production, trade, processing, and consumption at the forefront of any economic development strategy for the region” (Oerke *et al.*, 2012).

Farming is major source of income for many people in developing countries. In Ghana it represent 22% of the GDP (Ghana Statistical Service, 2014) and it is the main source of income to 60% of the population. The agricultural sector of the Ghanaian economy is dominated by peasant farmers who over the years continue to use traditional technology and farm inputs to produce crops. Although food crops such as maize, rice, plantain, cassava, yam, cocoyam and other vegetables are grown both on commercial and subsistence level all over Ghana, it is not sufficient enough to feed the Ghanaian populace as there is higher importation of cereals such as rice in the Country. The annual production in recent years continue to decline despite several programme and interventions by MOFA (Agyare *et al.*, 2014). However, as stated by Quansah in 2010 “increasing human population has led to intensive cultivation without adequately replenishing soil nutrients. The result has been the deterioration in crop yields and depletion of the resource base”. This decline is usually associated with the inputs cost and the fertility of the soil because crop production is strongly reliant on rainfall which varies due to climate changes.

"Ghana is in a unique position to not only leverage agriculture as an engine for poverty reduction and improved nutrition, but to become the breadbasket for the West Africa
countries in the sub-region” (Quansah, 2010). Ghana has an abundance of fertile land, water, and a generally favorable climate for agricultural production. However, in modern times the yields of agricultural crops especially cereal crops has not been encouraging. We therefore sort to assess the trend and significant differences in yields of major cereal crops across the ten (10) regions of Ghana over 2005 to 2014 period.

1.3. **Objective**

The main objective of the study was to analyze major cereal crop yields in Ghana over ten years period using MANOVA and Linear Mixed model.

The specific objectives were:

- To determine whether significant differences exist in annual maize yields in the ten regions of Ghana.
- To determine whether significant differences exist in annual rice yields in the ten regions of Ghana.
- To study the evolution of cereal crop yields between the regions in Ghana.

1.4. **Research Questions**

- Do differences exist in major cereal crops yields across the regions in Ghana?
- What is the trend of the major cereal crops yields in the country?

1.5. **Significance of the study**

The findings from this study would be contributing to the country’s agricultural sector in the following forms: First, it will guide farmers to know which cereal crop to invest in to increase yields. It will, again, enable the Ministry of Food and Agriculture (MOFA) to know which
cereal crop to intensify education on in order to increase yields across the districts and regions in Ghana.

Secondly, by examining yield differences among the districts, policy makers can be better informed and tailored to respond to challenges of food crop production among this important group of producers or farmers. The results will thus enable various stakeholders in Agriculture to increase their credit allocations to farmers.

Thirdly, the results will redirect stakeholders’ attention on the achievement of hunger target of the Sustainable Development Goal 2 (SDG 2) – on hunger, achieved food security and improved nutrition and promote sustainable agriculture by intensively increasing food crop production. Finally, the results would contribute to the existing knowledge and literature of academic sector. It will also be used as basis for research work in a related area or field of study.

1.6. Scope and Methodology

The study was carried out in Ghana. The data was obtained from Statistics, Research and Information Department (SRID) of Ministry of Food & Agriculture in Ghana. It comprised of crop yields data of two major cereal crops (Maize and Rice) in ten regions from 2005 to 2014. Crop yields were obtained from division of crop productivity in metric tons by the crop area in hectares (Mt/Ha). The data was then analyzed using MANOVA and Linear Mixed Models with aid of R software.

1.7. Organization of the study

The research is organized into five chapters. Chapter one is the introduction which comprises the background of the study, the motivation, objectives, research question, significant of the study, scope and methodology and organization. Chapter Two examines the related literature
to the research. The methods used for analysis of the data constitute Chapter three. Chapter four is the results and discussion of data and lastly chapter five presents conclusion and recommendations.
CHAPTER TWO

LITERATURE REVIEW

2.0. Introduction

This chapter explores related topics on how other researchers handled crop yields. It includes theoretical framework and conceptual topics such as mixed effects model of crop yields and modeling of crop yields in Ghana.

2.1. Theoretical Framework

“Major scientific research advancement has been considering food production as part of food security for quite a long time” (Ingram, 2011). “For instance in 1843 the Roth Amsted Experimental Station was established in UK, whilst the later part of the 19th century saw the swift growth of commercial plant breeding in Germany” (Harwood, 2005). Notwithstanding these many years’ research, there is still a need to institute how to produce more food given anticipated request (Godfray et al., 2010). “However, satisfying these increased demands poses huge challenges for the sustainability of both food production and aquatic ecosystems and the services they deliver to society” (Tilman et al., 2002).

Mindful of the adverse environmental concerns of most current food production techniques, it is clear that the necessary gains will have to be made in a more environmentally benign manner (Foresight, 2011; Gregory et al., 2002). “To this end, research has progressively concentrated on the production scheme in search of increasing efficiency by which inputs (especially nitrogen and water) are used, and reduce undesirable externalities such as water pollution, soil degradation, loss of biodiversity and conservatory gas emissions” (Gregory et al., 2002; Ittersum & Rabbinge, 1997).
Meanwhile, because the climate modification has intensified, research on food production has speedily increased. It is however clear that the change in climate will definitely affects crop growth in most part of the globe, with most harmful influence anticipated in the developing countries (Foresight, 2011; Parry et al., 2004; Parry, Rosenzweig & Livermore, 2005). In this regard reaching higher yields is part of the strategy for achieving food security while protecting the natural environment in developing Countries (Foley, 2011; Lobell, Cassman & Field, 2009).

Onumah and Aning (2009) in their “report specified that production of agricultural commodities in Ghana has recorded significant improvement since the late 1990s, with average growth in output rising between 2003-2006 to 5.8% from 3.6% in 1999-2002. He also noted that average yields for the major agricultural commodities have hardly risen since 2000, implying that the enhanced sector growth was chiefly due to strengthening of production. In addition, cereal crops such as maize, rice and millet, yam and cassava (starchy staples) amongst others have substantial potential for productivity improvements, which were 60 percent lower than attainable yields” (Onumah & Aning, 2009). These crops could increase yield over 60 percent if production technology is further improved.

“Major food and industrial commodities production is believed to have accounted for nearly 80% of Ghanaian’s GDP in the past. But this was stagnated as results of unimproved farm inputs used for production in the agricultural sector of the economy” (Onumah & Aning, 2009). Improving output and production in major crops aside cocoa will significantly reduce poverty amongst farmers who engage in non-cocoa growing crops especially in the northern part of Ghana. “Farm productivity has to be improved through measures such as modernizing the country’s agricultural marketing systems; aside the adoption of improved production technology such as planting high-yielding varieties and improved access to farm extension
and finance” (Onumah & Aning, 2009). Additional literary works by Denise Wolter in 2008 supports the assertion that Ghana’s current food production is low and must improve yields to fully enjoy the benefits of the commodities exchange.

Wolter affirms that Ghana continues to face problem with food security because of declined productivity in food crop sector and undeveloped local markets for most part of the country especially in northern part of Ghana (Wolter, 2008). In order to take advantage of the increasing demand of food in Ghana’s middle income economy, effort should be made to increase crop yields, improve the local food markets aside solving food security problems (Wolter, 2008). It is, however, disturbing how Ghana with favorable natural conditions for agricultural production but continues to rely heavily on imported food. “According to the Ministry of Food and Agriculture (MOFA), Ghana’s agricultural production currently meets only half of domestic cereal and meat needs and 60 per cent of domestic fish consumption (Wolter, 2008). Self-reliance is achieved only in starchy staples such as cassava, yam and plantain, while rice and maize production falls far below demand” (Wolter, 2008).

“The author further elaborated on various reasons why Ghana’s food crop production lies below potential yields. These factors in his research include untapped potential in irrigation; Ghana’s agriculture is mainly rain fed, with conventional systems of farming still dominant in most parts of the country (Wolter, 2008). Also, poor technology and small production units prohibits economies of scale and lead to sub-optimal yields. For instance, maize and rice are produced at a third of their potential yields per hectare due to poor technology and farm inputs” (OECD & FAO, 2009).

Also, inadequate transportation facilities and storage infrastructure to link farmers and market centers prevent small farmers to have access to internal and intentional markets. Currently, Ghana produces less than 30% of local based industry raw materials to process some of the
food crop (Wolter, 2008). “The Government of Ghana has implemented enticements such as tax holidays and subsidies to inspire food processing, but results have been minimal as holdups such as lack of infrastructure, and finance amongst others remain (OECD & FAO, 2009). As internal food processing remains below the potential to meet local demand, high-value food imports have been increasing” (Wolter, 2008).

In Diao, Hazell, and Thurlow (2010) paper on the “role of Agriculture in African development they discussed the yields of major crops in Ghana. According to them, the actual yields obtained are much lower than the achievable yields for many crops in most zones of Ghana. According to the Ministry of Food and Agriculture data they used, yields for most crops are 20%-60% below their achievable level under existing technologies combined with the use of modern inputs such as fertilizers and improved seeds, which provides an opportunity for agricultural growth” (Diao, Hazell & Thurlow, 2010).

2.2. Conceptual Framework

This section of the literature reviews the concepts and methods of other researchers that are related to this study. It includes topics such as; mixed effects model of crop yields, longitudinal study on crop yield and modeling for crop yields in Ghana.

2.2.1. A mixed effects model of crop yield

A study conducted by Verma et al. (2014) of department of mathematics & statistics, Haryana Agriculture University India, developed a methodology for pre-harvest crop yield prediction of major mustard growing districts in Haryana (India). They use linear mixed effects models with random time effects at district, zone and state level, to fit crop yield estimate. For mixed modeling incorporating the hierarchical data structure of yield was represented as

\[ y_{ijt} = s_i + z_{jt} + d_{ijt}, \]  \hspace{1cm} (2.1)
where

\[ y_{ijt} = \text{yield in the } j^{\text{th}} \text{ district within } i^{\text{th}} \text{ zone in the } t^{\text{th}} \text{ year,} \]

\[ s_i = \text{general state effect in the } t^{\text{th}} \text{ year,} \]

\[ d_{ijt} = \text{effect of the } j^{\text{th}} \text{ district within } i^{\text{th}} \text{ zone in the } t^{\text{th}} \text{ year,} \]

\[ z_{ij} = \text{effect of the } j^{\text{th}} \text{ zone in the } t^{\text{th}} \text{ year.} \]

For each of the three effects (state: \(s_i\), zone: \(z_{ij}\), district: \(d_{ijt}\)), they set up a time-series model with three components: Regression + Time Trend + White noise. Regression was a fixed part comprising regression on time as well as on the meteorological covariates. Time Trend comprised of a random part for serial correlation with covariance structure and regression splines. White noise was an additional independently distributed random error term.

The purpose of their study was to show the usefulness of the mixed model framework for pre-harvest crop yield forecasting. The findings indicated that there was improvement in the predictive accuracy of the zonal yield models using linear mixed modeling. The linear mixed models substantially improved the predictive accuracy and produced what they considered to be satisfactory district-level yield(s) estimation. They concluded by recommending the use of linear mixed models for pre-harvest yield forecasting of crop to enhance the predictive accuracy of the zonal models.

Roberts and Tack (2011) also presented a “paper on a mixed effects model of crop yields for purposes of premium determination at the Agricultural & Applied Economics Association at Pittsburgh, Pennsylvania, USA. The goal of their research was to determine empirical estimation of farm-level yield distributions, calculation of actuarially fair risk premiums, and prediction of potential efficiency gains using empirical mixed model for premium determination given below” (Robert & Tack, 2011).
where $y_{it}$ represents “yield on farm $i$ in year $t$, $a_i$ is a farm-specific intercept, $b_c$ is a county-specific trend, $c_i$ is a county-specific insurance effect, $\mu_{ct}$ is a county-specific random shock in year $t$, and $\varepsilon_{it}$ is an idiosyncratic random shock on farm $i$ in year $t$. The farm-specific intercept according to them can be written as a county-specific mean plus a farm-specific shock, $a_i = a_i + \mu_i$, and the county-specific slope parameters can be written as a population mean plus a county shock”, $b_c = b + \mu_{ct}$ and $c_c = c + \mu_{ct}$. (Robert & Tack, 2011).

These modifications generated the model

$$ y_{it} = a_c + bt + cins + \mu_i + \mu_{ct} + \mu_{c2} ins + \varepsilon_{it}. $$

(2.3)

They indicated two types of effects in this model. The first part $a_c + bt + cins$, represents the fixed effects of crop yields, and the second part, $\mu_i + \mu_{ct} + \mu_{c2} ins + \varepsilon_{it}$, represents random effects. They cited two random intercepts, farm-specific $\{\mu_i\}$ and county/time-specific $\{\mu_{ct}\}$; and, there were two random slopes at the county level, for the trend parameter $(\mu_{c1})$ and for the insurance parameter $(\mu_{c2})$.

“Followed conventional methods for estimating linear mixed models, they assumed that the random components follow a multivariate normal distribution and employed maximum likelihood estimation. With a separate model for each state-crop combination in their dataset, their empirical results suggested that there were several important sources of heterogeneity for crop yield distribution” (Robert & Tack, 2011).
For example, for Arkansas rice, the farm-specific random intercept accounted for 42% of the random variation of their model, the county-specific intercept accounted for 7.5%, the random slope for the time trend accounted for 0.1% and the random slope accounted for 1% (Roberts & Tack, 2011). These findings confirm their objectives that there were significant farm-level crop distribution shifters across space and time.

Another study conducted on longitudinal and spatial analyses applied to corn yield data from a long-term rotation trial by (Brownie, Larry & Tina, 1993) investigated a number of analyses of rotation trial of corn yields. The study used several types of split plot, repeated measures and spatial analyses, each with and without soil covariates and restricted attention to models that could be implemented using standard software. Important features of the yield data from these authors trial include considerable unbalance, possible correlations across time and space, possible heterogeneity in the error variances across years, and the availability of pretreatment measurements of soil properties.

Ignoring the soil covariates, a linear mixed model for the corn yields in their methodology was given as

\[
Y_{ijk} = \mu + \beta_i + (RY)_{jk} + \delta_{ij} + \epsilon_{ijk},
\]

where \( Y_{ijk} \) is the yield in year \( k \) for rotation \( j \) in block \( i \), \( k = 1,...,6,8,9,10, \) \( j = 1,...,27, i = 1,...,4, \) \( \beta_i \) is a random effect for the \( i^{th} \) block or rep, \( (RY)_{jk} \) is a fixed effect for rotation \( j \) in year \( k \), with \( \sum_{jk} (RY)_{jk} = 0, \)
\( \delta_{ij} \) is a random effect for the \( ij^{th} \) plot (the plot in block \( i \) containing rotation \( j \)), and \( \epsilon_{ijk} \) is a random error associated with the \( ij^{th} \) plot in year \( k \).
“The corn yield data were highly unbalanced because of the different crop sequences with corn planted in 155 of the possible 9 x 27 = 243 rotation-year combinations” (Brownie et al., 1993). “There were also several missing yields, thus instead of including main and interaction effects for rotation and year, they fitted an effect for each observed rotation-year combination, represented as \((RY)_j\) in the model above. The Authors’ diagnostic graphs and AIC values indicated that the model should allow for heterogeneity across years in the error variance. The precision of contrasts in years with high error variance according to their recommendation was underestimated in the analysis that assumed constant variance. Allowing spatial correlations resulted in a small reduction in estimates of precision, and including soil covariates which were important in two of the nine years used. The authors indicated that the best models were repeated measures with banded covariance, as well as autoregressive repeated measures and isotropic spatial models, both modified to allow heterogeneity of the error variance across years” (Brownie et al., 1993).

Also a study on climate variability and crop production in Tanzania conducted by Rowhani et al. (2011) revealed the relationship between seasonal climate and crop yields in Tanzania, focusing on maize, sorghum and rice. The study which makes use of linear mixed model outline below revealed an interesting result.

In order to determine the effects of climate on agricultural yields, and to exploit the cross-sectional and temporal attributes of their dataset, they developed linear mixed models for each of the three crops (maize, rice, and sorghum). The method was appropriate for longitudinal designs (Pinheiro et al., 2007; Zuur, 2010) where observations within a group are often more similar than would be predicted on a pooled-data basis. The model was given as

\[
y_{ij} = \beta_0 + \beta_1 T_{i,j} + \beta_2 P_{i,j} + \beta_3 P_{i,j}^2 + \beta_4 CW_{T,i,j} + \beta_5 CV_{P,i,j} + a_i + \epsilon_{ij},
\]

(2.5)
where $y_{ij}$ is yield, $i$ represents the regions and $j$ the observations within a region, $\beta_{0-5}$ represent model parameters, $a_i$ represents the random intercept term, $T$ represent temperature, $P$ represents precipitation and $\epsilon_{ij}$ is an error term. Also the model included a fixed part comprised of $P$ and $T$ (and their interaction term), $CV_T$ and $CV_P$ (and their interaction term), as well as $P^2$, and random intercepts.

“For comparison, the coefficients resulting from their mixed models were compared to those obtained from simple linear regression models that included the different regions as a dummy variable to account for fixed effects. Again a time variable (year from 1992) was also used in the linear models to capture yield changes related to non-climatic factors and other technological development” (Zuur, 2010).

$$y_j = \beta_0 + \beta_1 T_j + \beta_2 P_j + \beta_3 P^2_j + B_4 CV_{T-j} + B_5 CV_{P-j} + \beta_6 Region_j + \beta_7 Year_j + \epsilon_j. \quad (2.6)$$

“The linear models were developed using stepwise model selection based on the AIC. In order to compare climate data effects on yield estimates, they used both sets of models, the mixed and linear models, were also developed using climate data they extracted from the CRU dataset. The results of their study indicated that both intra and inter seasonal changes in temperature and precipitation influenced cereal yields in Tanzania. They projected that by their studies, in Tanzania, by 2050, projected seasonal temperature increases by 2°C will reduce average maize, sorghum and rice yields by 13%, 8.8% and 7.6% respectively” (Rowhani et al., 2011).

2.2.2. Modeling for crop yield in Ghana

Some researchers in Ghana have also conducted studies on crop yields using mixed modeling and other methods but usually concentrated on a single crop. For instance a study conducted
by Isaac (2012) of KNUST uses multiple comparison and random Effect model to analyze Cocoa production in Ghana from 1969/70 to 2010/11 production years.

In his studies, he set the linear mixed effects models as

\[ Y_i = X_i \beta + Z_i b_i + \epsilon_i, \quad i = 1, 2, \ldots, N, \]

(2.7)

where \( Y_i \) is a vector of observations with mean \( E(Y) = X_i \beta \)

\( X_i \) and \( Z_i \) are the design matrices corresponding to the fixed and random effects respectively, \( \beta \) is fixed effects vector, \( b_i \) is a vector of independent and identically distributed (iid) random effects with mean \( E(b_i) = 0 \) and variance-covariance matrix variable \( b_i \) = \( D \).

\( \epsilon_i \) is a vector of iid random errors with mean \( E(\epsilon_i) = 0 \) and variance \( \text{var}(\epsilon_i) = R \).

It was assumed that \( b_i \sim N(0, D) \) and \( \epsilon_i \sim N(0, \Lambda) \), with \( b_i \) independent of \( \epsilon_i \).

His analysis using mixed effect model, revealed that from 1969/70 production, all the six regions used in the study experienced increasing cocoa production trend with the exception of Volta region. Western and Ashanti regions had the highest production over the years.

Again, a study was carried out by Azinu (2014) of department of Crop Science University of Ghana, on the evaluation of hybrid maize varieties in three agro-ecological zones (transition forest, Guinea savannah and coastal savannah) in Ghana. The study which was undertaken to assess the relative yielding abilities and stability of 20 hybrids selected from the breeding programme of the West Africa Centre for Crop improvement of maize uses analysis of variance (ANOVA) model. ANOVA per location and across location or environment for agronomic traits were carried out using Genstat 12th edition.
Genotypes were considered as fixed effects, whilst environments and replication were considered as random effects. For each agronomic and morphological trait, an individual ANOVA was conducted by the researcher to determine the statistical significance of the genotypes at each environment and across environment. According to the author, the significant genotype-environment interaction revealed by additive main effect and multiplicative interaction (AMMI) analysis of variance for grain yield suggested that the relative performance of the genotypes changed for grain yield across all environments. The study also identified Wenchi as the location for the best grain yield performance and Tamale as an environment yielded low grain yield in both seasons.

2.2.3. The uniqueness of the study

The exploration of the related literature to our study has revealed some similarities and differences in our work. For instance the study by (Rowhani et al., 2011) in Tanzania uses LMM on three cereal crops (maize, rice and sorghum) with climate data to predict interseasonal variations in temperature and rainfall influenced on cereal yields in Tanzania similar to our study on two major cereal crops (maize and rice) in Ghana which also uses LMM and MANOVA to investigate significant differences and trend of the yields over time. However, our study does not consider climatic data and it was not for prediction purpose.

Unlike studies conducted by (Azinu, 2014; Isaac, 2012) in Ghana and considered only one crop at a time, our work is different in the sense that it considered two cereal crops at the same time. Also our study uses longitudinal data which most researchers in our terrain have not explored. Conducting research on crop yields again is a bit challenging due to missing values which might be present in the data setup and that was not different in our work. This usually occurred as results of unrecorded production by farmers. So being able to model two
most important cereal yields data in Ghana with our work prove the uniqueness of our study.

Our work might be among few studies conducted in this area of research in Ghana.
CHAPTER THREE

METHODOLOGY

3.0. Introduction

The chapter describes the models used and methods of analyzing the available data to satisfy the objectives of the study. Models under this discussion include MANOVA used to establish the differences in yields among the regions for each year and Linear mixed models used to assess the evolution of crop yields among the regions.

3.1. Data Description

The data used for this study is secondary data from the Department of Statistics, Research and Information of Ministry of Food & Agriculture (MOFA) in Ghana. It comprised of crop yields data of the two major cereal crops (maize and rice) in the ten (10) regions. Crop yield was defined by dividing crop productivity in metric tons with the crop area in hectares (Mt/Ha). Maize and Rice data were studied in each of the ten regions for the period 2005 to 2014. As at the time this study was carried out, maize and rice were the first and second most used cereals by Ghanaians. Also the current data on the two cereals for the ten years period available at MOFA was up to 2014 hence the period 2005 to 2014.

3.2. Multivariate Analysis Of Variance (MANOVA) Model

“MANOVA is an extension of analysis of variance (ANOVA) in which the effects of factors are assessed on a linear combination of several response variables. A multivariate generalization of the ANOVA-model was first addressed” by Wilks (1932). “MANOVA methodology is well established and widely used in many research areas ranging from Agriculture to psychology” (Casella & Berger, 2002; Zhang & Xiao, 2012). The model was
considered for our data because of the multivariate nature of our data and also because it usually give appropriate results for testing and comparing differences in means (yields) of crops (Zhang & Xiao, 2012).

3.2.1. One-way MANOVA model

In the one-way MANOVA-model, a single factor explains the variation in a set of response variables. The factor effect of one-way MANOVA-model for comparing g population mean is the following:

$$y_{i\ell} = \mu + \alpha_{i\ell} + \varepsilon_{i\ell}, \tag{3.1}$$

where $y_{i\ell}: p \times 1$ is vector of $p$ response variables for the $i^{th}$ replicate on the $\ell^{th}$ level of factor A, $\mu$ is the overall mean, $\alpha_i$ is group effect, $i=1,2,\ldots,n_i$, $\ell=1,2,\ldots,g$ and $\varepsilon_{i\ell}: p \times 1$ is the vector of effects for level $i$ of factor A. More specifically, $\alpha_i = \mu_i - \mu$ showing that the vector of effects could be interpreted as the deviation from the vector of overall means.

Further, it is assumed that errors are independently normally distributed with zero mean and constant covariance matrixes, $\varepsilon_{ik} \sim N_p(0, \Sigma)$ thus,

$$E(y_{i\ell}) = \mu + \alpha_i \text{ and } V(y_{i\ell}) = V(\varepsilon_{i\ell}) = \Sigma. \tag{3.2}$$

3.2.2. Two-way MANOVA with interactions

Similarly to the univariate model (3.1), the two-way MANOVA model with interactions which is considered in this thesis is expressed as:

$$y_{ij\ell} = \mu + \alpha_{j} + \beta_{ij} + \alpha_\beta_{ij\ell} + \varepsilon_{ij\ell}, \tag{3.3}$$
where \( y_{ij} : p \times 1 \) is vector of \( p \) response variables for the \( \ell \)th replicate on the \( i \)th level of factor A, and the \( j \)th level of factor B, \( \ell=1,2,\ldots,g, j=1,2,\ldots,b, i=1,2,\ldots,n \). In the two-way MANOVA-model, vectors \( \alpha_i, \beta_j \) and \( \alpha\beta_{ij} \) represent main and interaction effects respectively. Also, it is assumed that \( \epsilon_{ij\ell} \sim i.i.d. N_p (0, \Sigma) \) so that:

\[
E(y_{ij\ell}) = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} \quad \text{and} \quad V(y_{ij\ell}) = V(\epsilon_{ij\ell}) = \Sigma. \quad (3.4)
\]

We note that from (3.1) the model is over parameterized and so we impose the identifiability constraints \( \sum_{\ell=1}^{g} n_i \alpha_i = 0 \),

that is, all group effects add up to zero and there are \( n_i \) individuals in each group.

The following terminology is sometimes useful:

1. Systematic part: \( \mu + \mu_{\ell} \)
2. Random part: \( \epsilon_{i\ell} \)

The observations can be decomposed as

\[
x_{ij\ell} = \bar{x} + (x_{i\ell} - \bar{x}) + (x_{ij\ell} - x_{i\ell}) \quad (3.5)
\]

That is,

Observation is equal to overall mean + treatment effect (or group effect) + residuals

This implies

\[
(x_{ij\ell} - \bar{x})(x_{ij\ell} - \bar{x}) = [(x_{i\ell} - \bar{x}) + (x_{ij\ell} - x_{i\ell})][(x_{i\ell} - \bar{x}) + (x_{ij\ell} - x_{i\ell})]
\]
\[ \begin{align*}
&= (\bar{x}_i-x)(\bar{x}_i-x)' + (\bar{x}_i-x)(\bar{x}_i-x)' \\
&+ (x_i-x_i)(\bar{x}_i-x)' + (x_i-x_i)(\bar{x}_i-x)'
\end{align*} \]

Summing over \( i \), we obtain the following;

\[ \sum_{i=1}^{n}(x_i-x)(x_i-x)' = \sum_{i=1}^{n}(\bar{x}_i-x)(\bar{x}_i-x)' + (\bar{x}_i-x)\sum_{i=1}^{n}(x_i-x_i)' \]

\[ + \sum_{i=1}^{n}(x_i-x_i)(x_i-x_i)' + \sum_{i=1}^{n}(x_i-x_i)(x_i-x_i)' \]

\[ = n_s(\bar{x}_i-x)(\bar{x}_i-x)' + \sum_{i=1}^{n}(x_i-x)(x_i-x)' \]  
Since \( \sum_{i=1}^{n}(x_i-x_i) = 0 \).

Next, we sum over \( \ell \) and obtain;

\[ \sum_{\ell=1}^{g}\sum_{i=1}^{n}(x_{i\ell}-\bar{x}_{i\ell})(x_{i\ell}-\bar{x}_{i\ell})' = \]

\[ \sum_{\ell=1}^{g}n_{i\ell}(\bar{x}_{i\ell}-\bar{x}_{i})(\bar{x}_{i\ell}-\bar{x}_{i})' + \sum_{\ell=1}^{g}\sum_{i=1}^{n}(x_{i\ell}-\bar{x}_{i})(x_{i\ell}-\bar{x}_{i})' \]

where \( T: p \times p \) is the total SSCP matrix, \( B \) is the “between” matrix which is denoted by \( H \), the “hypothesis” and “within” matrix \( W \) is often denoted as \( E \), “error” matrix.

The within matrix is only meaningful if \( \Sigma \) is constant over samples.
Certainly;

\[
E = \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)
\]

\[
= (n_i - 1)S_1 + (n_2 - 1)S_2 + \ldots + (n_g - 1)S_g 
\]  

(3.7)

3.2.3. Hypothesis testing in the MANOVA model

“The testing of hypothesis for MANOVA-model is based on the partitioning of sums of squares becomes more complex because of the interrelationships between the \(p\) include ANOVA-models. Unlike the univariate models, one must now consider sums of squares but also across products for the factors in the MANOVA-model. In the resulting matrices, called sums of squares and cross products (SSCP), diagonal elements corresponds to the usual sums of squares for each of the \(p\) response variables whereas the off-diagonal elements correspond to the cross products for each response variable pair” (Zhang & Xiao, 2012).

When data is balanced, the partitioning of SSCP matrices is independent in analogy with the ANOVA-models described earlier. For instance, in the MANOVA-model:

\[
T = H + E 
\]  

(3.8)

where \(T: p \times p\) is the total SSCP matrix, \(H:p \times p\) is the hypothesis SSCP matrix and \(E:p \times p\) is the error SSCP matrix. The general hypothesis under the MANOVA can be written as:

\[
H_0: \mu_1 = \mu_2 = \ldots = \mu_g 
\]  

(3.9)

\(H_i\) : at least one group centroid is different,

with \(p\) responses, \(1\) factor and \(g\) groups. The assumptions are given as

1. Independent groups, independent observations
2. Responses are multivariate normal with each group
3. Population covariance matrices are equal across groups

3.2.4. Construction of Test Statistics for MANOVA-model

From (3.6), we recall that;

\[ H = \sum_{i=1}^{g} n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^t \]
\[ E = \sum_{i=1}^{g} \sum_{i=1}^{n_i} (x_{i,i} - \bar{x})(x_{i,i} - \bar{x})^t \]
\[ H+E = \sum_{i=1}^{g} \sum_{i=1}^{n_i} (x_{i,i} - \bar{x})(x_{i,i} - \bar{x})^t \]

The simplest way to characterize information contain in positive semi-definite matrices is through eigenvalues \((\lambda, \theta, \phi)\).

Now we consider the root equations

\[ |E - \lambda (H+E)| = 0 \]
\[ |H - \theta (H+E)| = 0 \text{ and} \]
\[ |H - \phi E| = 0 \] .

Thus, the eigenvalues of matrices

\[ E(H+E)^{-1}, \ H(H+E)^{-1}, \ E^{-1} H. \]

Indeed, \[ |H - \phi E| = 0 \Leftrightarrow |E^{-1} H - \phi I| = 0. \]

Now rephrasing the first and the second root equation, we obtain;
\[(1-\lambda)E - \lambda H = 0\]

\[|\mu - \frac{1-\lambda}{\lambda} E| = 0\]

\[|H - \phi E| = 0\]

and

\[(1-\theta)H - \theta E = 0\]

\[|H - \frac{\theta}{1-\theta} E| = 0\]

Thus

\[\phi = \frac{1-\lambda}{\lambda} = \frac{\theta}{1-\theta}, \quad \theta = \frac{\phi}{1+\phi} = 1 - \lambda\]

\[\lambda = \frac{1}{1+\phi} = 1 - \theta.\]

Johnson and Wichern (2007) outline four (4) statistics commonly used in MANOVA-model and considered in this study as

1. **Wilks’ Lambda(\(\Lambda\)):**

Under the null hypothesis of no factor \(\gamma\) effects:

\[H_0 : \gamma = 0,\]

where \(\gamma\) is average yields of the regions.

The likelihood ratio test statistic is given as
\[ \Lambda = \frac{|E|}{|E + H|} = \det(E(H + E)^{-1}) \]

\[
= \prod_{j=1}^{p} \lambda_j = \prod_{j=1}^{p} \theta_j = \prod_{j=1}^{p} \frac{1}{1 + \phi_j}
\] (3.12)

is generally known as Wilks’ \( \Lambda \) after Wilks (1932). The null hypothesis is rejected for small values of \( \Lambda \), showing that \( E \) is small compared to the total SSCP matrix \( E + H \).

2. Hotelling-Lawley Trace

The test statistic is given by:

\[
U = tr(HE^{-1}) = \sum_{j=1}^{p} \theta_j = \sum_{j=1}^{p} \frac{\theta_j}{1 - \theta_j}
\] (3.13)

is often referred to as Hotelling-Lawley Trace after Lawley (1938) and Hotelling (1951) who took part in developing the statistic. Naturally, a large \( H \) relative to \( E \) would indicate a larger support for \( H \) and a larger trace. Hence the null hypothesis (3.11) of no effects is rejected for large values of \( U \).

3. Pillai’s Trace:

Pillai (1955) developed the following statistic:

\[
V = tr((E + H)^{-1}H) = \sum_{j=1}^{p} \theta_j
\] (3.14)

which is commonly known as Pillai’s Trace. As with Hotelling-Lawleys Trace, the null hypothesis is rejected for large values of \( V \), indicating a large \( H \) relative to \( E \) (Khatri & Pillai, 1968).
4. Roy’s Greatest Root:

Roy (1953) also developed the Statistic:

The largest root of the equation $|H - \phi E| = 0$

$$\phi_{max} = \frac{\theta_{max}}{1 - \theta_{max}},$$  \hspace{1cm} (3.15)

which is commonly known as Roy’s Greatest Root is also one of the statistic usually used.

“Wilk’s Lambda is use for measuring amount of variability not explained by the levels of the independent variables. Pillai’s Trace is best for general use especially when the sample size is uneven and data is correlated. Hotelling-Lawley Trace is generally converted to Hotellings T-square which is used when the independent variables forms two groups and represents the most significant linear combination of the dependent variable. Roys greatest root or eigenvalue is generally preferred if the outcome variables are highly inter-correlated reflecting single construct” (Harrar, 2008; Khatri, 1966).

“Wilks’ Lambda, Hotelling-Lawley Trace, Pillai’s Trace and Roy’s Greatest Root are exact tests, meaning that the probability of rejecting $H_0$ in (3.11) is equals to $\alpha$” (Rencher, 2003).

“However, these tests have different probabilities of rejecting $H_0$ when it is false, thus implying that the tests have different power for a given sample. In general none of these multivariate tests is uniformly better than the other, although there might be situations where one test is preferred (Harrar & Bathke, 2008; Littell, Stroup & Freund, 2002). Also for large samples, all these tests are essentially equivalent” (Johnson & Wichern, 2007).
3.2.5. Assumptions of MANOVA model

The following assumptions as applied to MANOVA were discussed by Casella and Berger (2002) have been outlined here.

**Normal Distribution:** “The dependent variable should be normally distributed within groups. Overall, the $F$ test is robust to non-normality, if the non-normality is caused by skewness rather than by outliers. Tests for outliers should be run before performing a MANOVA, and outliers should be transformed or removed. Normal probability plot is used to test for normality. Here each residual is plotted against its expected value under normality. A plot that is nearly linear suggests agreement with normality, whereas a plot that departs substantially from linearity suggests that the error distribution is not normal” (Casella & Berger, 2002).

**Linearity:** “MANOVA assumes that there are linear relationships among all pairs of dependent variables, all pairs of covariates, and all dependent variable-covariate pairs in each cell. Therefore, when the relationship deviates from linearity, the power of the analysis will be compromised. This can be checked by conducting scatter plot matrix between the residuals of the dependent variables” (Casella & Berger, 2002). A good fit for scatter plot should not produce a pattern in the plot. Linearity should be met for each group of the MANOVA separately.

**Homogeneity of Variances:** “Homogeneity of variances assumes that the dependent variables exhibit equal levels of variance across the range of predictor variables” (Littell et al., 2002). We recollect that the error variance is computed (SS error) by adding up the sums of squares within each group. “If the variances in the two groups are different from each other, then adding the two together is not appropriate, and will not yield an estimate of the common within-group variance. Homoscedasticity can be examined graphically or by means
of a number of statistical tests such as Bartlett’s and Brown Forsyth Test. The significant of the p-value ($p < 0.05$) for both tests indicates that variances are not identical” (Littell et al., 2002).

**Homogeneity of Covariance**: Equality of covariance matrices is an assumption checked by running a Box’s M test. Unlike most tests, the Box’s M test tends to be very strict, and thus the level of significance is typically 0.001. So as long as the $p$ value for the test is above 0.001, the assumption is met (Box, 1949).

Thus, with Box’s M, one is interested in testing the null hypothesis:

$$H_0 : \Sigma_1 = \Sigma_2 = \ldots = \Sigma_i = \ldots = \Sigma_k = \Sigma,$$  \hspace{1cm} (3.16)

where $\Sigma_i$: $p \times p$ is the covariance matrix of the $i^{th}$ combination of factor, $i = 1, 2, \ldots k$, in the MANOVA-model with $p$ response variables. Setting $n = \sum_{i=1}^{k} n_i$ and $\nu_i = n_i - 1$, under the null hypothesis (3.16), the pooled estimator of the total covariance matrix is:

$$S = \sum_{i=1}^{k} \frac{\nu_i S_i}{n - k},$$ \hspace{1cm} (3.17)

where $n_i$ is the number of replicates on the $i^{th}$ factor combination, and $S_i$ is unbiased estimator of $\Sigma$. A generalized likelihood ratio test statistic can then be calculated as

$$M = (n - k) \log |S| - \sum_{i=1}^{k} \nu_i \log S_i.$$ \hspace{1cm} (3.18)

Using scale factors, Box’s M could be approximated to either a $\chi^2$ or a $F$-distribution. For both approximations, the null hypothesis of homoscedasticity is rejected for large values of the scaled test statistics (Box, 1949). As Timm (2002) notes, the $\chi^2$ approximation is
preferred when \( n < 20, p < 6 \) and \( k < 6 \). Otherwise, \( F \) approximation is recommended. “Box’s M test unlike most tests tends to be very strict, and thus the level of significance is typically 0.001. So as long as the \( p \) value for the test is above 0.001, the assumption is met” (Box, 1949).

**Multicollinearity:** “This assumption occurred when the correlations among the independent variables is so high to be accepted. When this happens the effects of independents cannot be separated. Under this assumption, the estimates are unbiased but the explanatory variables and their joint effects assessment are not reliable. As a rule of thumb, when inter-correlation among independent variables are above 0.80, it signals a possible problem of multicollinearity” (Garson, 2012).

### 3.2.6. Advantage of MANOVA model

The MANOVA-model has many advantages over simultaneous estimation of several ANOVA-models;

MANOVA tests whether there are significant differences among the means of factor levels on several dependent variables. Thus using MANOVA, one is able to test joint hypotheses of all univariate ANOVA models and more likely to observe differences between factor levels. For instance, two factors may have no main or interaction effects on two different response variables separately but only jointly (Sawyer, 2009).

“Fitting one MANOVA-model instead of several ANOVA-models decreases the experiment with Type I error probability. As a simple example, suppose that \( \alpha = 5\% \) for \( F \)-tests in 6 separate ANOVA-models. Then, the experiment type I error would be equal to 30\% whereas an overall \( F \)-test for models in the MANOVA-model would imply a 5\% Type I error probability” (Littell *et al.*, 2002).
Several ANOVA-models estimated individually does not take into account the co-variance pattern among dependent variables. On the other hand, the MANOVA model is useful not only to mean differences of factor levels but also to the covariation between response variables. When response variables are studied together, they are likely to be correlated to at least some extend and by conducting several ANOVA analyses this correlation would be lost (Littell et al., 2002).

3.2.7. Limitations of MANOVA model

Outliers: Like ANOVA, MANOVA is extremely sensitive to outliers. Outliers may produce either a Type I or Type II error and give no indication as to which type of error is occurring in the analysis. There are several programs available to test for univariate and multivariate outliers. Tests for outliers should be run before performing a MANOVA, and outliers should be transformed or removed.

Multicollinearity and Singularity: “When there is high correlation between dependent variables, one dependent variable becomes a near-linear combination of the other dependent variables. Under such circumstances, it would become statistically redundant and suspect to include both combinations. Absence of multicollinearity is checked by conducting correlations among the dependent variables. The dependent variables should all be moderately related, but any correlation over 0.80 presents is a concern for multicollinearity” (Zhang & Xiao, 2012).

“Another setback of MANOVA as compared to univariate ANOVA-model, is the difficulty of the MANOVA-model as the number of factors included in the model hastily increases. The model specification in many ways similar to its univariate analogues presented earlier. The assumptions for the MANOVA-model are the same as for the ANOVA-model, but extended to comprise multivariate normality” (Littell et al., 2002).
Also equality of covariance matrices for factor combinations is assumed that

\[ \Sigma_{11} = \Sigma_{12} = \cdots = \Sigma_{1b} = \cdots = \Sigma_{ab} = \Sigma, \]

(3.19)

where \( \Sigma: p \times p \) is an unknown covariance matrix.

### 3.3. Linear Mixed Model (LMM)

Another model used in the study is Linear mixed model which is an extension of Linear model for data that were collected and summarized in groups (profile) over a period (Faraway, 2005; Johnson & Wichern, 2007; West, Welch & Galecki, 2014). The model is used in outcome variables in which the residuals are normally distributed but may not be independent or have constant variance. We used LMM because our data is longitudinal or repeated-measures in which subjects (yields) are measured repeatedly over time. LMM may include both fixed effects and random effects parameters.

#### 3.3.1. The model formulation

Unlike linear model where \( E(Y) = X_{} \beta \) with \( \beta \) as fixed effects; in LMM, \( X_{} \beta \) is used for the fixed effects and \( b_i \) used as random effects. The model is specified conditionally as

\[ E(Y/b) = X_{} \beta + Z_{} b_i \]

(3.20)

where \( b_i \) is the conditional mean and realization of the random variable. The distribution assumptions are:

\[ b_i \sim N(0, D) \text{ and } \xi_i \sim N(0, \Sigma). \]

(3.21)

The random components \( b_i \) and \( \xi_i \) are independent.

Also
\[ \text{var}(Y/b) = \Sigma \text{ and } Y \sim (X\beta, ZDZ + \Sigma). \] 

(3.22)

The parameters of this model are:

Fixed –effects \( \beta \) and random effect components, unknowns in the variance matrices \( D \) and \( \Sigma \).

The unknown variance elements are also referred to as the covariance parameters and collected in the vector \( \theta \). The vector of covariance parameters,

\[ \theta = \begin{bmatrix} \theta_D \\ \theta_Z \end{bmatrix}, \] 

(3.23)

combines all parameters from the covariance matrices \( D \) and \( \Sigma \) in the vectors \( \theta_D \) and \( \theta_Z \) respectively.

### 3.3.2. Assumptions of LMM

Linear Mixed model has similar assumptions as linear model but in the case of LMM, the data might not necessary have constant variance although it might be normally distributed and linear. This is because of the presence of correlation within independent variables. The assumptions are given as

1. \( E(\varepsilon) = 0 \) for all \( i = 1, 2, ..., n \) or, equivalently, \( E(Y_i) = \beta_0 + \beta_i X_i \).

2. \( \text{Var}(\varepsilon_i) = \sigma^2 \) for all \( i = 1, 2, ..., n \) or, equivalently, \( \text{var}(Y_i) = \sigma^2 \).

3. \( \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \) for all \( i \neq j \), or equivalently, \( \text{cov}(Y_i, Y_j) = 0 \).

The first assumption is that \( Y_i \) depends only on \( X_i \) and that all other variation in \( Y_i \) is random.

The second assumption asserts that the variance of \( \varepsilon \) on \( Y \) does not depend on the value of \( Y \).

Assumption two is also known as the assumption of homoscedasticity, homogeneous
variance or constant variance. Under the third assumption, the \( E \) variables (or the \( Y \) variables) are uncorrelated with each other (Rencher & Schaalje, 2008).

### 3.3.3. The implied marginal model

The linear mixed effects model (3.20) implies the marginal linear model

\[
Y_i = X_i \beta + \xi_i^*,
\]

(3.24)

where \( \xi_i^* = Z_i \beta_i + \xi_i \).

Thus, the \( \xi_i^* \) is normally distributed with expected value

\[
E(\xi_i^*) = E(Z_i \beta_i) + E(\xi_i)
\]

\[
= Z_i E(\beta_i) + E(\xi_i)
\]

\[
= Z_i 0 + 0 = 0
\]

and covariance matrix

\[
\text{Cov}(\xi_i^*) = \text{Cov}(Z_i \beta_i) + \text{Cov}(\xi_i)
\]

\[
= Z_i \text{Cov}(\beta_i) Z_i^T + \text{Cov}(\xi_i)
\]

\[
= Z_i D Z_i^T + \Sigma_i.
\]

By defining the marginal variance-covariance matrix as

\[
V_i = Z_i D Z_i^T + \Sigma_i,
\]

(3.25)

we get \( \xi_i^* \sim N_{m_i}(0, V_i) \).
Hence, the marginal distribution of $Y_i$ is defined as

$$Y_i - N(X_i \beta, V_i).$$  \hspace{1cm} (3.26)

3.4. Estimation of model parameters

The most commonly used estimation of model parameters is done by maximum likelihood estimation (MLE) and restricted maximum likelihood estimation (REML). Both estimation procedures were employed in this study.

3.4.1. Maximum Likelihood Estimation (MLE)

This requires maximizing the implied marginal likelihood. Bruce Schaalje (2008); (Rencher & Schaalje, 2008) developed the likelihood by specifying the likelihood contribution from measurement $Y_i$ conditioned on the random effects $b_i$.

The marginal distribution of $Y_i$ (3.24), is the multivariate normal density function

$$f(Y|\beta, \theta) = (2\pi)^{-n/2} \text{det}(V_i)^{-1/2} \exp \left(-\frac{1}{2}(Y_i - X_i \beta)^T V_i^{-1} (Y_i - X_i \beta) \right),$$

where $\text{det}$ is the determinant and $V_i$ is given by (3.25).

Hence, given the observed data $Y_i = y_i$, the likelihood function contribution for the $i^{th}$ subject is $L_i(\beta, \theta) = (2\pi)^{-n/2} \text{det}(V_i)^{-1/2} \exp \left(-\frac{1}{2}(y_i - X_i \beta)^T V_i^{-1} (y_i - X_i \beta) \right)$.

We have observed $m$ independent subjects and the product of these $m$ likelihood functions, gives us the joint likelihood function

$$L(\beta, \theta) = \prod_{i=1}^{m} L_i(\beta, \theta)$$
\[
= \prod_{i=1}^{n} (2\pi)^{-\frac{n}{2}} \det(V_i)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y_i - X_i\beta)^T V_i^{-1}(y_i - X_i\beta)\right).
\]

Hence, the log-likelihood function is

\[
l(\beta, \theta) = -\frac{1}{2} n \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{n} \ln(\det(V_i)) - \frac{1}{2} \sum_{i=1}^{n}(y_i - X_i\beta)^T V_i^{-1}(y_i - X_i\beta).
\]

By assuming that \( \theta \) is known, the log-likelihood function becomes a function of \( \beta \) only. This leads to the maximization of the log-likelihood function (3.27) being equivalent to the minimization of its last term

\[
q(\beta) = -\frac{1}{2} \sum_{i=1}^{n}(y_i - X_i\beta)^T V_i^{-1}(y_i - X_i\beta).
\]

By using the method of generalized least squares, we minimize \( q(\beta) \) to find \( \hat{\beta} \).

\[
\frac{\partial q(\beta)}{\partial \beta} = \frac{\partial}{\partial \beta} \left[-\frac{1}{2} \sum_{i=1}^{n}(y_i^T V_i^{-1}X_i\beta - \beta^T X_i^T V_i^{-1}y_i + \beta^T X_i^T V_i^{-1}X_i\beta)\right] = 0
\]

\[
\Rightarrow \sum_{i=1}^{n}(-X_i^T V_i^{-1}y_i + X_i^T V_i^{-1}X_i\beta) = 0
\]

\[
\Rightarrow \hat{\beta} = (\sum_{i=1}^{n} X_i^T V_i^{-1}X_i)^{-1} \sum_{i=1}^{n} X_i^T V_i^{-1}y_i.
\]

\( \hat{\beta} \) can be written as \( b^T Y \), which is the best linear unbiased estimator (BLUE) of \( \beta \) which means that \( E[b^T Y] = \beta \) and that it has the smallest variance among all unbiased linear estimators.
3.4.2. Restricted Maximum Likelihood Estimation (REML)

The REML estimation is an alternative way of estimating the covariance parameters in $\theta$, which is often preferred to ML estimation due to the fact that it produces unbiased estimates of covariance parameters by taking into account the degrees of freedom that were lost as a results from estimating the linear fixed effects in $\beta$ (Bruce Schaalje, 2008).

The REML log-likelihood function is given by

$$ l_{REML}(\theta) = -\frac{1}{2} (n-p) \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{m} \ln(\text{det}(V_i)) $$

(3.29)

$$ = -\frac{1}{2} \sum_{i=1}^{m} (r_i^T V_i^{-1} r_i) - \frac{1}{2} \sum_{i=1}^{m} \ln(\text{det}(X_i^T V_i^{-1} X_i)) $$

(3.30)

where $r_i$ is given as $r_i = y_i - X_i \hat{\beta} = y_i - X_i ((\sum_{i=1}^{m} X_i^T V_i^{-1} X_i)^{-1} \sum_{i=1}^{m} X_i^T V_i^{-1} y_i)$.

Here we observe that the difference between the ML and the REML log-likelihood function is that the REML subtracts less in the first term, $n-p$ and an extra term $\frac{1}{2} \sum_{i=1}^{m} \ln(\text{det}(X_i^T V_i^{-1} X_i))$. The general motivation for using REML is to obtain unbiased estimates of the covariance parameters.

3.5. Information Criteria

The study used two types of information criteria often used to choose the best fitted model for the data, the Akaike information criteria and the Bayes information criteria developed by (Akaike, 1981) and (Burnham & Anderson, 2004) also known as Schwarz Criterion (Schwarz, 1978), respectively.

The Akaike information criteria, $AIC$, is defined by
\[ AIC = -2l(\hat{\beta}, \hat{\theta}) + 2p \]  

(3.31)

where \( l(\beta, \theta) \) can be either the ML or REML log-likelihood function and \( p \) represents the total number of parameters, both the fixed and random effects, being estimated in the model. The model with the lowest \( AIC \) value is assumed to be the best fit for the data.

The Bayes information criteria, \( BIC \) is defined by

\[ BIC = -2l(\hat{\beta}, \hat{\theta}) + p \ln(n) \]  

(3.32)

where \( l(\beta, \theta) \) is the ML log-likelihood function, \( p \) represents the total number of parameters, both the fixed and random effects, being estimated in the model and \( n \) is the total number of observations used in estimation of the model. That is \( n = \sum_{m=1}^{m} n_i \).

According to Bates and Pinheiro (2006) we can calculate the REML version of the BIC by simply using REML log-likelihood function and replacing \( \ln(n) \) by \( \ln(n - p_{\text{fixed}}) \), where \( p_{\text{fixed}} \) is the number of estimated fixed effect parameter in the model, in Equation (3.32). In other words, the \( BIC \) applies a greater penalty for models with more parameters than the \( AIC \). And as such, the lowest \( BIC \) value is assumed to be the best fit model for the data and preferable for the data. West et al. (2014) cited that there is no information criterion which stands apart as the best criterion to be used when selecting linear mixed effects models.

### 3.6. Diagnostics

It is necessary after a linear mixed effects model is fitted to check whether the underlying distributional assumptions for the random effects and the residuals appear valid for the data. Diagnostic methods for linear models are well established, but diagnostics for linear mixed
effects models are, however, more difficult to perform and interpret due to the complexity of the model.

“The most useful method for diagnostics according to Bates and Pinheiro (2006) are based on plots of the residuals, the fitted values and the estimated random effects”. In this thesis, we will do diagnostic by using the functions `qqnorm.lme`, `qqline`, `histogram` and `plot.lme` in (Pinheiro et al., 2011). Here, the standardized residuals, defined as the raw residuals divided by the estimated corresponding standard deviation, are used.

3.7. **Statistical software used**

MANOVA and Linear mixed procedures in the Statistical Software (R) were used in fitting the models since it allows for numerical integration of the normal random effects by Gaussian and adaptive Gaussian quadrature methods (Bates & Pinheiro, 2006; Jose & Bates, 2000). The conditional likelihood can be fed to the program as well to complete the full marginalization. The SPSS and Excel statistical software were considered for data processing and manipulations.

3.8. **Compound Symmetry (CS)**

The structure of compound symmetric specifies that observations on the same subject have homogeneous covariance and homogeneous variance. The variance of the response variable is equal to variance of between and within subjects (Littell, Pendergast & Natarajan, 2000).

Thus, $\text{cov}(Y_{ijt}, Y_{ijk}) = \sigma_{CS,b}^2$ if $t \neq k$, $V(Y_{ijk}) = \sigma_{CS,b}^2 + \sigma_{CS,w}^2$ \hspace{1cm} (3.33)

where $Y_{ijt}$ is the value of the response valuable (yield) measured at year $t$ and $k$ on region $j$ in district $i$, $i = 1, \ldots, g$ and $j=1, \ldots, n_{10}$, $t= 1, \ldots, 10.$
Also the covariance \( \text{cov}(Y_{ij}, Y_{jk}) = \sigma_{i,k} \) is the covariance measures at year \( t \) and \( k \) on the same subject (yield) and \( \sigma_{i} = \sigma_{i}^{2} \) is the variance at time \( t \). The function of the correlation of CS is given as \( \text{corr}_{CS}(\text{yield}) = \frac{\sigma_{CS,b}^{2}}{\sigma_{CS,b}^{2} + \sigma_{CS,w}^{2}} \) \( \text{(3.34)} \)

We notice that the correlation does not depend on the value of the yield, in the sense that the correlation between two observations is equal to all pairs of observations on the same region. The structure of Compound symmetry is occasionally called 'variance components' structure since the two parameters \( \sigma_{CS,b}^{2} \) and \( \sigma_{CS,w}^{2} \) denotes between-subjects and within-subject variances. The structure of compound symmetry is given as

\[
\Sigma = \begin{bmatrix}
\sigma_{1}^{2} & \sigma_{1} & \cdots & \sigma_{1} \\
\sigma_{1} & \sigma_{1}^{2} & \cdots & \sigma_{1} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{1} & \sigma_{1} & \cdots & \sigma_{1}^{2}
\end{bmatrix}, \text{ thus } \Sigma = \begin{bmatrix}
1 & \rho & \cdots & \rho \\
\rho & 1 & \cdots & \rho \\
\vdots & \vdots & \ddots & \vdots \\
\rho & \rho & \cdots & 1
\end{bmatrix}
\]
CHAPTER FOUR

DATA ANALYSIS AND DISCUSSION OF RESULTS

4.0. Introduction

This various analyses of the study and the discussion of the results obtained are presented in this chapter.

4.1. Descriptive analysis

This thesis applies MANOVA and Linear Mixed models to analyze cereal crop yields in Ghana. All the analysis in this study is limited to the yields of the two major cereal crops (Maize and Rice) recorded over the period 2005 to 2014. The unit of measurement is the 2016 districts in the country. The previous section’s methodologies are applied to the data and the results are discussed in this section.

Table 4.1 present descriptive analyses of the results. The lowest average maize yields for the period under study was 0.13 metric tons per hectare (Mt/Ha), this was recorded in 2006 at Bongo in Upper East region and the maximum average maize yields was 9.56 Mt/Ha and that occurred in 2012 at Dormaa East District in Brong Ahafo region. In the same period the lowest average rice yield recorded was 0.68 Mt/Ha in 2011 at Bolga in Upper East region. The maximum average rice yields for the ten years period was 6.72 Mt/Ha and that was recorded in 2005 at Dangme West District in Greater Accra region. We observed from Table 4.1 that the average yield of rice in the period is higher than the average maize yield. This indicates that rice yields in Ghana for the period under study are on average higher than maize yields in most of the regions for the ten years period. Figure 4.1 give graphical representation of this results.
Table 4.1: Descriptive statistics of the two major cereal crops

<table>
<thead>
<tr>
<th>Yield</th>
<th>Minimum</th>
<th>1st Quartile</th>
<th>3rd Quartile</th>
<th>Mean</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize</td>
<td>0.130</td>
<td>1.252</td>
<td>1.900</td>
<td>1.605</td>
<td>9.560</td>
</tr>
<tr>
<td>Rice</td>
<td>0.680</td>
<td>1.265</td>
<td>2.665</td>
<td>2.098</td>
<td>6.720</td>
</tr>
</tbody>
</table>

Source: SRID of MOFA

Figure 4.1: Line graph showing average yields by year

4.2. Preliminary analysis

Analysis of box plot appears to indicate steady improvement in the yields of rice over the years especially between 2009 and 2014. However, the yields of maize within the same period seem to be declining. Figure 4.2 demonstrate this assessment.
4.2.1 Mean plots over time

The average yields of maize and rice for the ten regions is presented in the mean profile in Figure 4.3 and 4.4. We observed that from 2005 to 2006, the yields of maize decrease slowly over the years for all the ten (10) regions. The yields increased in 2008, 2010 and 2012, and then decreased in 2009, 2011, 2013 and 2014. The overall yields of maize as can be seen from Figure 4.3 indicates unstable trend in all the regions.

Figure 4.3: Mean profile of maize yields
Unlike maize yields which were decreased steadily within the period, rice yields, however, increased gradually from 2005 to 2009 with sharp increased between 2009 and 2010. Then it continues progressively until 2014 as indicated in Figure 4.4. The progressive trend in the rice yields from 2009 might be as results of National Rice Development Strategy (NRDS) implemented in 2009 by MOFA which increased domestic production of rice between 2009 and 2014 from 235,000Mt to 417,000Mt (GhanaWeb, 2015).

Figure 4.4: Mean profile of rice yields

4.3. MANOVA Analysis

4.3.1. Test on Assumptions

Diagnostics were fitted to check whether the underlying distributional assumptions for MANOVA model appear valid. These assumptions were as follows;

**Linearity:** Linearity assumption was assessed by scatter plot for both maize and rice yields. The result is indicated in Figure 4.9 in appendix A. The plot suggests that response variables were linear as it does not show patterns in the plot.

**Normality:** Box plot for both maize and rice yields, Q-Q normality plot and histogram with density curve in Figure 4.8 and 4.9 in appendix A seem to suggest that the error points have
not deviated from the fitted line for both maize and rice yields. A formal test was conducted using Shapiro-Wilk test to test for normality. From Table 4.6c in Appendix A, the Shapiro-Wilk test confirmed that the residuals are normally distributed since p-value of 0.00713 is less than significance level of 0.05 ($\alpha = 0.05$).

**Homogeneity of Variance:** Bartlett's test and Brown Forsyth test on the yields in Tables 4.6d and 4.7 also in appendix A indicates that variance were not identical as the p-values of both tests were less than $\alpha = 0.05$.

**Outliers:** From Table 4.6c in Appendix A, the Bonferroni test obtained a p-value of 0.0420 which is less than the significance level (0.05) suggesting the presence of outliers. The most extreme observations in both yields were removed from the data set.

**Homogeneity of Covariance:** Box M-test in also in Table 4.6c in Appendix A confirmed that equality of covariance assumption was not met as the p-value 2.2e-16 is less than 0.001 indicating significant of M which also implies that groups differ from each other.

**Multicollinearity:** The assumption was tested by inter-correlation among the dependent variables. As indicated in Table 4.8a and b in appendix A, none of the correlation is above 0.80 signifying the absent of multicollinearity effects.

4.3.2. MANOVA Analysis

The results of multivariate analysis of variance conducted on the data set using the four most commonly used statistic discussed in section 3.2.5 revealed that, significant differences exist in the yields of two major cereal crops in the country. For instance, the results of Pillai’s test in Table 4.2a indicate that the yields of the two cereal crops are significantly different in across the various regions. All the statistics has P-value less than 0.05, signifying significant evidence against the null hypothesis that, there are no significant differences in the average yields of the two major cereal crops across the regions. The analysis of which crop have
significant differences in Table 4.2b indicates that differences occurred in both maize and rice yields across the regions. Further analysis in Table 4.3 on the mean vectors revealed significant differences in the yields between the regions.

Table 4.2a: MANOVA Test for Differences in crop yields across the ten regions

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>Estimates</th>
<th>DF</th>
<th>Approx. F</th>
<th>Pr (&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Region</td>
<td>Residual</td>
</tr>
<tr>
<td>Pillai's trace</td>
<td>0.76033</td>
<td>9</td>
<td>1185</td>
<td>80.756</td>
</tr>
<tr>
<td>Wilks' lambda</td>
<td>3.75e-01</td>
<td>9</td>
<td>1185</td>
<td>83.164</td>
</tr>
<tr>
<td>Hotellin-Lawley</td>
<td>1.30240</td>
<td>9</td>
<td>1185</td>
<td>85.596</td>
</tr>
<tr>
<td>Roy's greatest root</td>
<td>0.90123</td>
<td>9</td>
<td>1185</td>
<td>118.66</td>
</tr>
</tbody>
</table>

***significant at 5%

Table 4.2b: Test of Significant difference in crop yields

| Dependent Variables | Value  | F     | Hypothesis df | Error df | Pr (>F).
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MAIZE AND RICE YIELDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pillai's Trace</td>
<td>0.458</td>
<td>4.595</td>
<td>178</td>
<td>2754</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>0.586</td>
<td>4.731</td>
<td>178</td>
<td>2752</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>0.63</td>
<td>4.868</td>
<td>178</td>
<td>2750</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>0.469</td>
<td>7.261</td>
<td>89</td>
<td>1377</td>
</tr>
<tr>
<td>MAIZE YIELDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pillai's Trace</td>
<td>0.185</td>
<td>3.521</td>
<td>89</td>
<td>1377</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>0.815</td>
<td>3.521</td>
<td>89</td>
<td>1377</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>0.228</td>
<td>3.521</td>
<td>89</td>
<td>1377</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>0.228</td>
<td>3.521</td>
<td>89</td>
<td>1377</td>
</tr>
<tr>
<td>RICE YIELDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pillai's Trace</td>
<td>0.298</td>
<td>6.561</td>
<td>89</td>
<td>1377</td>
</tr>
<tr>
<td>Wilks' Lambda</td>
<td>0.702</td>
<td>6.561</td>
<td>89</td>
<td>1377</td>
</tr>
<tr>
<td>Hotelling's Trace</td>
<td>0.424</td>
<td>6.561</td>
<td>89</td>
<td>1377</td>
</tr>
<tr>
<td>Roy's Largest Root</td>
<td>0.424</td>
<td>6.561</td>
<td>89</td>
<td>1377</td>
</tr>
</tbody>
</table>

***significant at 5%

Pairwise comparisons on the mean vectors of the yields were further performed to determine where the differences occurred across the regions. The results of the test for both maize and
rice yields are presented in Table 4.3 for five regions and Table 4.16a in appendix E for all the regions. Table 4.3 indicates that within the ten years study period, significant differences in both maize and rice yields occurred in most of the regions average yields. This is noted by the significant level less than 0.05 \((\text{sig.} < 0.05)\) of the regions’ pairwise comparisons with other regions. For instance, differences occurred in the comparisons between average maize yields of Ashanti region and six of the regions. However, no differences occurred in the average yields of Ashanti region and Upper East, Upper West and Western regions. Similarly, significant differences occurred in the average rice yields of Ashanti regions and all the regions except Northern region.

Also pairwise comparisons of average maize yields of Brong Ahafo region and other regions indicates that significant differences occurred between the region and seven other regions’ yields except the yields of Central and Eastern regions. In the same way, significant differences occurred in the average rice yields of Brong Ahafo region and seven other regions with exception of Central and Upper West regions where there were no significant differences. Details of other regions results as well as 95% confidence interval for both yields are present in Table 4.3 and Table 4.16a of appendix E.
Table 4.3: Pairwise Comparisons of differences in mean yields

<table>
<thead>
<tr>
<th>(I) Region</th>
<th>(J) Region</th>
<th>Differences in Mean yields (I-J)</th>
<th>Std. Error</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maize</td>
<td>Rice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASHANTI</td>
<td>B/AHAFO</td>
<td>-.613*</td>
<td>.449*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>CENTRAL</td>
<td>-.548*</td>
<td>.565*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>EASTERN</td>
<td>-.701*</td>
<td>-.350*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>G/ACCRA</td>
<td>.280*</td>
<td>-1.400*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>NORTHERN</td>
<td>-.130*</td>
<td>.090</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>UPPER EAST</td>
<td>-0.009</td>
<td>-.469*</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>UPPER WEST</td>
<td>.13</td>
<td>.684*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>VOLTA</td>
<td>-.287*</td>
<td>.620*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>WESTERN</td>
<td>-0.035</td>
<td>.902*</td>
<td>1.0000</td>
</tr>
<tr>
<td>B/AHAFO</td>
<td>ASHANTI</td>
<td>.613*</td>
<td>-.449*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>CENTRAL</td>
<td>.065</td>
<td>.115</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>EASTERN</td>
<td>-.088</td>
<td>-.800*</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>G/ACCRA</td>
<td>.893*</td>
<td>-1.850*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>NORTHERN</td>
<td>.483*</td>
<td>-.359*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>UPPER EAST</td>
<td>.604*</td>
<td>-.919*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>UPPER WEST</td>
<td>.483*</td>
<td>.235</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>VOLTA</td>
<td>.326*</td>
<td>-1.069*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>WESTERN</td>
<td>.578*</td>
<td>.452*</td>
<td>0.0000</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>ASHANTI</td>
<td>.548*</td>
<td>-.565*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>B/AHAFO</td>
<td>-.065</td>
<td>-.115</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>EASTERN</td>
<td>-.153</td>
<td>-.915*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>G/ACCRA</td>
<td>.827*</td>
<td>-1.965*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>NORTHERN</td>
<td>.418*</td>
<td>-.475*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>UPPER EAST</td>
<td>.538*</td>
<td>-1.034*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>UPPER WEST</td>
<td>.418*</td>
<td>.0119</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>VOLTA</td>
<td>.260*</td>
<td>-1.185*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>WESTERN</td>
<td>.513*</td>
<td>.337*</td>
<td>0.0000</td>
</tr>
<tr>
<td>EASTERN</td>
<td>ASHANTI</td>
<td>.701*</td>
<td>-.350*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>B/AHAFO</td>
<td>.088</td>
<td>.800*</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>CENTRAL</td>
<td>.153*</td>
<td>-.915*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>G/ACCRA</td>
<td>.980*</td>
<td>-1.050*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>NORTHERN</td>
<td>.571*</td>
<td>.440*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>UPPER EAST</td>
<td>.691*</td>
<td>-.119</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>UPPER WEST</td>
<td>.571*</td>
<td>1.034*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>VOLTA</td>
<td>.413*</td>
<td>-.270*</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>WESTERN</td>
<td>.666*</td>
<td>1.252*</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

*significant at 5%
Another statistical test that was used to determine which regions differ significantly from others was Bonferroni Post hoc comparisons on the regions yields for both maize and rice yields. This test as indicated in Table 4.16b and 4.16c of appendix E also revealed similar results to the Pairwise Comparisons of differences in mean yields.

4.4. LMM analysis

This section of the analysis deals with the linear mixed model considered for the study. The model was based on the methodology in 3.3. Two types of mixed models were considered for the study: model with random intercepts and model with random intercepts and slope.

4.4.1. The Model Specification

The model with random intercept is given as

$$ y_{ij} = \beta_0 + \sum_{r=1}^{q} \beta_r X_{ij} + \beta_{1i} Year_i + \sum_{r=1}^{q} \beta_r X_{ij} Year_i + b_{0i} + \xi_{ij} \quad (4.1) $$

$\beta_0$ is the intercept of the fixed effects, $\sum_{r=1}^{q} \beta_r X_{ij} + \beta_{1i} Year_i$,

where $y_{ij}$ is a $(n \times 1)$ vector of the $j^{th}$ region yields for the $i^{th}$ year,

$\sum_{r=1}^{q} \beta_r X_{ij} Year_i$ is the interaction between regions and years,

$X_{ij}$ is $j^{th}$ region of year $i$,

$\beta_r$ is a $(p \times 1)$ vector of fixed effects parameters, that is region

$b_{0i}$ is $(g \times 1)$ vector of random effects that occur in the data vector $y_{ij}$ and

$\xi_{ij}$ is an $(n \times 1)$ vector of errors (also random) specific to $j^{th}$ region of year $i$. 
The model with random intercept and slope is given as

\[ y_{ij} = \beta_0 + \sum_{r=1}^{g} \beta_r X_{ij} + \beta_{0r} Year_i + \sum_{r=1}^{g} \beta_r X_{ij} Year_i + b_{0i} + b_{1i} Year + \xi_{ij} \]

(4.2)

\[ z_g = (1 | Year) \]

\[ b = (b_{0i}, b_{1i}) \]

where \( b_{0i} \) is the random intercept and \( b_{1i} \) is the random slope for year, all other variables retain their usual meaning.

In order to know which model best fit the data, we used Akaike Information Criteria (AIC) and the Bayes Information Criteria (BIC) developed by (Akaike, 1981) and (Burnham & Anderson, 2004) to select best fitted model to the data. Table 4.4 presents the two models and the values of AIC and BIC. It is important to state that for all analyses, P-value < 0.05 was considered to be statistically significant.

**Table 4.4: Selection of best model**

<table>
<thead>
<tr>
<th>Maize</th>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>Test</th>
<th>L.Ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random intercept</td>
<td>1</td>
<td>22</td>
<td>1326.305</td>
<td>1442.632</td>
<td>-641.1527</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ran. int. and slope</td>
<td>2</td>
<td>24</td>
<td>1329.980</td>
<td>1456.882</td>
<td>-640.9902</td>
<td>1 vs 2</td>
<td>0.3250162</td>
<td>0.85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rice</th>
<th>Model</th>
<th>df</th>
<th>AIC</th>
<th>BIC</th>
<th>logLik</th>
<th>Test</th>
<th>L.Ratio</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random intercept</td>
<td>1</td>
<td>22</td>
<td>3084.077</td>
<td>3200.403</td>
<td>-1520.039</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ran.int. and slope</td>
<td>2</td>
<td>24</td>
<td>3087.685</td>
<td>3214.587</td>
<td>-1519.843</td>
<td>1 vs 2</td>
<td>0.3917895</td>
<td>0.821</td>
</tr>
</tbody>
</table>

From Table 4.4, we observed that the p-value associated with random intercept and slope model of maize is not statistically significant (0.85> 0.05) which indicates random intercept model best fit the data than random intercept and slope. It is also confirmed by the lower AIC and BIC values associated with the model (1326.305 and 1442.632).
The rice model also revealed similar findings. It also indicated that the model with random intercept best fit the data than random intercept and slope. As it can be seen from the Table 4.4, the p-value is not statistically significant (0.821>0.05) for random intercept and slope which was supported by lower AIC and BIC values given by the model (3084.077 and 3200.403).

4.4.2. Diagnostics on the fitted LMM model

Further assessment on the fitted LMM for both maize and rice were conducted. The plots were (1) QQ-plot, (2) scattered plot and (3) a histogram with a density curve of the fitted LMM. A good fit for QQ-plot should produce a straight one-to-one line of points, for the scatter plot, a good fit should not produce a pattern in the plot. Generally, QQ-plot is often desired to the scatter plot. The normality nature of the density curve on the histogram with the points and approximated linearity of the QQ-plot indicates that, the model is an effective model. Hence we conclude that the diagnostic plots support the fitted model and so LMM is a good model for fitting maize and rice yields. This is clearly shown in Figure 4.5.
Figure 4.5: Diagnostic plots for fitted Linear Mixed model
4.4.3. Random effects estimates

The first part of the output that we would like to discuss is the random effect estimates. The random effects (random intercepts) are random values associated with the levels of a random factor (District) in the LMM. These values which are specific to a given level of a random factor District represent random deviations from a given district from overall fixed effects as represented in Table 4.5.

Table 4.5: Random effects estimates

<table>
<thead>
<tr>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
<th>Groups</th>
<th>Name</th>
<th>Variance</th>
<th>Std.Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DISTRICT (Intercept)</td>
<td>0.5885</td>
<td>0.2426</td>
<td></td>
<td>DISTRICT (Intercept)</td>
<td>0.1911</td>
<td>0.4371</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>0.1314</td>
<td>0.3212</td>
<td></td>
<td>Residual</td>
<td>0.2668</td>
<td>0.5165</td>
<td></td>
</tr>
<tr>
<td>95% Conf. Interval</td>
<td>Lower 0.220</td>
<td>Est. 0.251</td>
<td>Upper 0.285</td>
<td>95% Conf. Interval</td>
<td>Lower 0.542</td>
<td>Est. 0.609</td>
<td>Upper 0.685</td>
</tr>
</tbody>
</table>

The standard deviation column measures the variability for each random effect added to the model. As we can see from Table 4.5, maize has much less within variability (24.26%) than rice (43.71%). The variance column shows the variability of the intercept (Districts) across the regions. We observed that our model indicates 58.85% of variability in maize yields and 19.11% of rice yields across the ten regions of Ghana.

Of course, not all the variability was accounted for by our model and that is indicated in the residuals which stand for variability within districts. This value is 32.12% and 51.65% for maize and rice yields respectively. It is believed that there is other variability from unknown covariates which also contribute to deviations in yields of these cereal crops. The graphical visualization of the distributions of districts as intercepts is given below. Clearly, Figure 4.6 shows that maize yields has average less within variability than rice yields.


4.4.4. Fixed effects parameter estimates

This show a portion of the model output that contain the fixed effect parameter estimates, their corresponding standard errors, the degrees of freedom, the t-test values and the corresponding p-values as well as their interactions. The intercept estimate which is also the reference category (Ashanti region) gives the expected average yields in metric tons per hectare when all the predictors or covariates are zero. From Table 4.6a, our model indicated that on the average, the expected maize yield is 1.37Mt/Ha if all other regions recorded zero yields.

The estimate column tell us that Brong Ahafo region have average yields of maize higher than Ashanti region. This is indicated by the positive coefficient (0.71) figure of the region’s estimate. Again, the estimates column indicated that Central region have average yields of maize lower than Ashanti region for the study period as indicated by the negative sign of the coefficient (-0.49) of the region’s estimate. The rest of the regions estimates in Table 4.6a are interpreted in similar way.

The most important aspect of the fixed effects analysis is the interaction between the regions and the year. The interaction terms can be interpreted as value telling us the relationship between the estimated change in regional maize yields and the year. Thus, interaction
between the two gives suggestions in the model as what trend is found in the maize yields for the period under study.

As a form of demonstration, suppose the interaction between two regions and year give the model, \( Y = \beta_0 + \beta_1 R_1 + \beta_2 R_2 + \beta_3 Year + \beta_4 R_1 Year + \beta_5 R_2 Year \) (4.3)

where \( Y \) is the response variable (yields), \( \beta_0 \) is intercept, \( \beta_i \) is the coefficient parameter, \( R_i \) is region 1 and \( R_j \) is region 2, \( R_k \) is the reference region.

Model for \( R_1 \)

\[
Y_{R_1} = \beta_0 + \beta_1 + \beta_3 Year + \beta_4 Year
\]

\[
= (\beta_0 + \beta_1) + (\beta_3 + \beta_4) Year
\]

\[
\Rightarrow \alpha_0 + \alpha_1 Year. \quad (4.4)
\]

Model for \( R_2 \)

\[
Y_{R_2} = \beta_0 + \beta_2 + \beta_3 Year + \beta_4 Year
\]

\[
= (\beta_0 + \beta_2) + (\beta_3 + \beta_4) Year
\]

\[
\Rightarrow \alpha_0 + \alpha_1 Year. \quad (4.5)
\]

Model for \( R_3 \)

\[
Y_{R_3} = \beta_0 + \beta_5 Year \quad (4.6)
\]

Now the difference between region 1 and 2 and reference region model gives

\[
Y_{R_1} - Y_{R_0} = \beta_1 + \beta_4 Year
\]

\[
Y_{R_2} - Y_{R_0} = \beta_2 + \beta_4 Year \quad (4.7)
\]

This is graphically presented as
Thus, the difference between the interaction of regions and the year determine the trend in the model. The results of these interactions between the regions and year for both rice and maize yield models are also given in Table 4.6a and 4.6b respectively.

Table 4.6a results suggest a trend in maize yields of six regions that is decelerating as indicated by the negative (-) sign attached to the estimated values. Only three of the regions (Brong Ahafo, Northern and Western) aside Ashanti region has positive interaction average fixed effect values which mean that they have progressive trend in maize yields for the study period.
Table 4.6a: Fixed effects parameter estimates for maize yields

<table>
<thead>
<tr>
<th>Maize</th>
<th>Estimates</th>
<th>Std.error</th>
<th>df</th>
<th>t-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(INTERCEPT)</td>
<td>1.37177</td>
<td>0.07036</td>
<td>1231</td>
<td>19.495406</td>
<td>0.0000***</td>
</tr>
<tr>
<td>BRONG. AHAFO</td>
<td>0.71144</td>
<td>0.10177</td>
<td>231</td>
<td>6.900433</td>
<td>0.0000***</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>-0.49031</td>
<td>0.11528</td>
<td>231</td>
<td>-4.253051</td>
<td>0.0000***</td>
</tr>
<tr>
<td>EASTERN</td>
<td>-0.30681</td>
<td>0.10129</td>
<td>231</td>
<td>-3.029021</td>
<td>0.0027***</td>
</tr>
<tr>
<td>G. ACCRA</td>
<td>-0.51641</td>
<td>0.12798</td>
<td>231</td>
<td>-4.035011</td>
<td>0.0001***</td>
</tr>
<tr>
<td>NORTHERN</td>
<td>0.01693</td>
<td>0.12956</td>
<td>231</td>
<td>0.168178</td>
<td>0.8666</td>
</tr>
<tr>
<td>UPPER EAST</td>
<td>-0.36260</td>
<td>0.10068</td>
<td>231</td>
<td>-3.053178</td>
<td>0.0025***</td>
</tr>
<tr>
<td>UPPER WEST</td>
<td>-0.14901</td>
<td>0.10997</td>
<td>231</td>
<td>-1.149953</td>
<td>0.2514</td>
</tr>
<tr>
<td>VOLTA</td>
<td>-0.06050</td>
<td>0.11342</td>
<td>231</td>
<td>0.550207</td>
<td>0.5827</td>
</tr>
<tr>
<td>WESTERN</td>
<td>0.02617</td>
<td>0.07079</td>
<td>1231</td>
<td>0.569605</td>
<td>0.569</td>
</tr>
<tr>
<td>YEAR</td>
<td>0.04511</td>
<td>0.00797</td>
<td>1231</td>
<td>0.569605</td>
<td>0.569</td>
</tr>
<tr>
<td>B. AHAFO:YEAR</td>
<td>0.01065</td>
<td>0.01141</td>
<td>1231</td>
<td>0.933553</td>
<td>0.0357***</td>
</tr>
<tr>
<td>CENTRAL:YEAR</td>
<td>-0.00680</td>
<td>0.01339</td>
<td>1231</td>
<td>-0.507377</td>
<td>0.0010***</td>
</tr>
<tr>
<td>EASTERN:YEAR</td>
<td>-0.06290</td>
<td>0.01184</td>
<td>1231</td>
<td>-5.317655</td>
<td>0.0000***</td>
</tr>
<tr>
<td>G. ACCRA:YEAR</td>
<td>-0.04810</td>
<td>0.01461</td>
<td>1231</td>
<td>-3.293161</td>
<td>0.0000***</td>
</tr>
<tr>
<td>NORTHERN:YEAR</td>
<td>0.01296</td>
<td>0.01142</td>
<td>1231</td>
<td>1.135008</td>
<td>0.0006***</td>
</tr>
<tr>
<td>UPPER EAST: YEAR</td>
<td>-0.05130</td>
<td>0.01490</td>
<td>1231</td>
<td>-3.441122</td>
<td>0.2566</td>
</tr>
<tr>
<td>UPPER WEST:YEAR</td>
<td>-0.04080</td>
<td>0.01462</td>
<td>1231</td>
<td>-2.787415</td>
<td>0.0054***</td>
</tr>
<tr>
<td>VOLTA:YEAR</td>
<td>-0.03590</td>
<td>0.01241</td>
<td>1231</td>
<td>-2.889813</td>
<td>0.0039***</td>
</tr>
<tr>
<td>WESTERN:YEAR</td>
<td>0.00415</td>
<td>0.01332</td>
<td>1231</td>
<td>0.311693</td>
<td>0.7553</td>
</tr>
</tbody>
</table>

***significant at 5%

Table 4.6b present fixed effects parameter estimates of rice yields. Similar to the maize yields, the intercept (Ashanti region) is the expected average yields of rice in metric tons per hectare when all covariates are zero. Our model indicates that the expected average yields of rice decreases by 1.035Mt/Ha if other regions recorded zero yields.

The estimate column tell us that Brong Ahafo region have average yields of rice higher than Ashanti region. This is indicated by the positive coefficient (0.49) figure of the region’s estimate. Also the estimates column for Central region indicates that it have average yields of rice higher than Ashanti region for the study period as given by the positive coefficient (0.32) of the region’s estimate. The rest of the regions estimates in Table 4.6a revealed that all the regions have average rice yields higher than Ashanti region for the study period.
The interaction between the regions and year gives significant positive estimates suggesting a trend in rice yields that is progressively increasing for each of the regions. This confirms the earlier assessment in Figure 4.2 which indicated stable improvement in rice yields for all the regions.

Table 4.6b: Fixed effects Parameter estimates for rice yields

<table>
<thead>
<tr>
<th>Region</th>
<th>Estimates</th>
<th>Std.error</th>
<th>df</th>
<th>t-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERCEPT</td>
<td>-1.0349446</td>
<td>0.14509882</td>
<td>1231</td>
<td>-7.132688</td>
<td>0.0000***</td>
</tr>
<tr>
<td>BRONG. AHAFO</td>
<td>0.4862897</td>
<td>0.21185939</td>
<td>231</td>
<td>2.295342</td>
<td>0.0226***</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>0.3202962</td>
<td>0.23532551</td>
<td>231</td>
<td>1.361077</td>
<td>0.0048***</td>
</tr>
<tr>
<td>EASTERN</td>
<td>0.5833881</td>
<td>0.20834345</td>
<td>231</td>
<td>2.800126</td>
<td>0.0055***</td>
</tr>
<tr>
<td>G. ACCRA</td>
<td>0.8772819</td>
<td>0.26037494</td>
<td>231</td>
<td>3.369303</td>
<td>0.0009***</td>
</tr>
<tr>
<td>NORTHERN</td>
<td>1.0327226</td>
<td>0.20976651</td>
<td>231</td>
<td>4.923201</td>
<td>0.0000***</td>
</tr>
<tr>
<td>UPPER EAST</td>
<td>1.5809686</td>
<td>0.24136448</td>
<td>231</td>
<td>6.550131</td>
<td>0.0000***</td>
</tr>
<tr>
<td>UPPER WEST</td>
<td>0.1551816</td>
<td>0.27012671</td>
<td>231</td>
<td>0.574477</td>
<td>0.0062***</td>
</tr>
<tr>
<td>VOLTA</td>
<td>0.7140924</td>
<td>0.22778364</td>
<td>231</td>
<td>3.134959</td>
<td>0.3763</td>
</tr>
<tr>
<td>WESTERN</td>
<td>0.2051626</td>
<td>0.23146262</td>
<td>231</td>
<td>0.886375</td>
<td>0.0019***</td>
</tr>
<tr>
<td>YEAR</td>
<td>0.1867735</td>
<td>0.01420502</td>
<td>1231</td>
<td>13.148423</td>
<td>0.0000***</td>
</tr>
<tr>
<td>B.AHAFO:YEAR</td>
<td>0.1538969</td>
<td>0.02024638</td>
<td>1231</td>
<td>7.601207</td>
<td>0.0000***</td>
</tr>
<tr>
<td>CENTRAL:YEAR</td>
<td>0.1284677</td>
<td>0.02409227</td>
<td>1231</td>
<td>5.332321</td>
<td>0.0000***</td>
</tr>
<tr>
<td>EASTERN:YEAR</td>
<td>0.0624697</td>
<td>0.02112496</td>
<td>1231</td>
<td>2.957152</td>
<td>0.0032***</td>
</tr>
<tr>
<td>G.ACCRA:YEAR</td>
<td>0.0739272</td>
<td>0.02582751</td>
<td>1231</td>
<td>2.862344</td>
<td>0.0043***</td>
</tr>
<tr>
<td>NORTHERN:YEAR</td>
<td>0.1908038</td>
<td>0.02022598</td>
<td>1231</td>
<td>9.433598</td>
<td>0.0000***</td>
</tr>
<tr>
<td>UPPER EAST:YEAR</td>
<td>0.1679122</td>
<td>0.02653888</td>
<td>1231</td>
<td>6.327024</td>
<td>0.0000***</td>
</tr>
<tr>
<td>UPPER WEST:YEAR</td>
<td>0.1647742</td>
<td>0.02581505</td>
<td>1231</td>
<td>6.382877</td>
<td>0.0000***</td>
</tr>
<tr>
<td>VOLTA:YEAR</td>
<td>0.0199702</td>
<td>0.02210353</td>
<td>1231</td>
<td>0.903483</td>
<td>0.3664</td>
</tr>
<tr>
<td>WESTERN:YEAR</td>
<td>0.1900362</td>
<td>0.02394832</td>
<td>1231</td>
<td>7.935261</td>
<td>0.0000***</td>
</tr>
</tbody>
</table>

***significant at 5%

Approximately 95% confidence intervals were obtained for the parameters. These confidence intervals are based on the asymptotic normality of the covariance parameter estimates. Since most of the confidence intervals do not contain zero, it confirms our findings for both fixed and random effects discussed in 4.3.1 and 4.3.2. For approximated 95% confidence intervals for fixed effects see Table 4.14a and b in appendix C.
4.5. Discussion

Research has indicated that modeling longitudinal data over time with Linear Mixed Model is an appropriate method (Brownie, 1993; Bruce & Schaalje, 2008). The result of our study revealed interesting findings about maize and rice yields in the country for the period under study. One such finding was the successive improvement of rice yields from 2009 to 2014 for eight of the ten regions in Ghana. Further research indicated that the implementation of National Rice Development Strategy (NRDS) accounted for this improvement. This was in consistent with a study done by Isaac (2012) of KNUST on Cocoa production for six regions of Ghana where he noticed an increasing trend of cocoa production for five of the six regions considered for the study.

Unlike the rice yields, the trend in maize yields was generally decreasing especially for the last four years (2010-2014). This decreasing trend calls for reform on the production of maize crop. Against the popular believe that maize is the most produced and consumed cereal crop in Ghana and therefore will have higher yields, the study did indicated that rice yields are averagely higher than maize yields in the country. Also, it revealed that significant differences exist in both maize and rice yields among all the regions in Ghana and there is variability between and within the yields among the regions with maize having much less within variability than the rice yields. The study identified that significant differences occurred in the majority of the regions’ yields with other regions for both maize and rice. Only few of the regions indicated absent of differences between their yields and that of other regions.

Although, the study did show that modeling cereal yields over time with MANOVA and LMM is valid, we think that modeling more crops with variables such as rainfall and climate data together with the Districts covariates will reveal more interesting findings. This could be done by joint model where more crops could be used. There is the need therefore, to focus...
more on production of maize crop in order to reverse the decreasing trend in the yields of maize. This means that MOFA should increase intensive education on the crop with much more viable seeds and incentives to farmers who engage in production of the crop. The most important aspect of our study is the fact that it has revealed the need to do further investigation on cereal crop yields in the country with other methods such as joint models. With this analysis, we will be able to determine the significant level of major crops produced and consume in Ghana as well as the evolution among the crops.
CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.0. Introduction

This is the last chapter of this thesis which presents the conclusion and recommendation based on the results of the study.

5.1. Conclusion

This study uses data from Statistics Research and Information Department of Ministry of Food and Agriculture (MOFA) to analyze major cereal crop yields in Ghana. The study was aimed at investigating whether significant differences exist in the crop yields and to determine the evolution of crop yields between the regions in Ghana. MANOVA and Linear Mixed Models (LMM) were used to assess these objectives in two major cereal crops (Maize and Rice) yields produced and consumed in Ghana. The Restricted Maximum Likelihood Method was employed to select best fitted LMM to estimates random and fixed effects parameters. The diagnostic plots indicate that, the LMM selected is good for determine the evolution of cereal crop yields.

It was found that significant differences exist in maize and rice yields across the ten (10) regions in Ghana. The study identified that the differences occurred in the majority of the regions yields for both maize and rice with only few of them indicating absent of differences in their yields comparisons.

Descriptive statistics revealed that within the ten years period considered for the study, Brong Ahafo region recorded the highest average maize yields of 9.56 metric tons per hectare (Mt/Ha) and the lowest average maize and rice yields was recorded in Upper East region.
(0.13Mt/Ha and 0.68Mt/Ha). The highest average rice yield was recorded in Greater Accra region (6.72 Mt/Ha).

The study also confirms that there is variability in cereal crop yields of maize and rice between and within the regions with maize yields having much less within variability than rice yields. It reveals a trend that is decelerating in maize yields of majority of the regions and steadily increasing trend in rice yields in all the regions. Against the popular believe that maize is number one cereal crop produced and consumed in Ghana and hence may have higher average yields, the study generally indicated that, the yields of rice are averagely higher than the yields of maize from 2005 to 2014.

5.2. **Recommendation**

Based on these findings, maize production should be given more attention such as a national programme to improve the yields so as to reverse the declining trend likewise rice production which has been given a boost due to NRDS programme. This will improve yields of maize and increase credit facility to farmers as well as reducing poverty and hunger.

MOFA is also encouraged to intensify education on the two cereals widely consume in Ghana and give more benefits to farmers who engage in its production, especially the youth.

We also recommend that researchers engaging in similar work of study should strive to obtain data on rainfall and climate in addition to yields data. This will clarify more significant differences in cereal yields in the various regions in Ghana and help to determine more interesting patterns.

We therefore conclude by declaring that Joint Models that will include other factors such as rainfall and climate data which may influence crop yields should be employed to investigate cereal crop yields in Ghana.


Quansah, G. W. (2010). *Effect of organic and inorganic fertilizers and their combinations on the growth and yield of maize in the semi-deciduous forest zone of Ghana.* Department of Crop and Soil Sciences, College of Agriculture and Natural Resources, Kwame Nkrumah University of Science and Technology, Kumasi.


APPENDIX

Appendix A: Test of Assumptions

Figure 4.7. Scatter plot of maize and rice yields

Figure 4.8. Boxplot of maize and rice yields

Figure 4.9. Test Normality assumption
Table 4.6c: Test of Assumptions

**Bonferroni Test:** \( r_{student} = 3.6472 \) Unadjusted p-value = 0.0005 Bonferroni p-value = 0.0402

**Shapiro-Wilk normality test:** \( W = 0.9889 \), p-value = 0.007226

**Box's M-test for Homogeneity of Covariance Matrices**

\( \text{Chi-Sq (approx.)} = 140.943 \), df = 20, p-value < 2.2e-16
Table 4.6d: Bartlett's Test of Homogeneity of variance

<table>
<thead>
<tr>
<th>YEAR</th>
<th>Bartlett's K-square</th>
<th>P-value</th>
<th>Bartlett's K-square</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize</td>
<td>Rice</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>92.794</td>
<td>214.526</td>
<td>9</td>
<td>4.47e-16</td>
</tr>
<tr>
<td>2006</td>
<td>61.815</td>
<td>76.054</td>
<td>9</td>
<td>5.98e-10</td>
</tr>
<tr>
<td>2007</td>
<td>50.814</td>
<td>137.409</td>
<td>9</td>
<td>7.57e-08</td>
</tr>
<tr>
<td>2008</td>
<td>97.238</td>
<td>93.671</td>
<td>9</td>
<td>2.20e-16</td>
</tr>
<tr>
<td>2009</td>
<td>108.028</td>
<td>114.713</td>
<td>9</td>
<td>2.20e-16</td>
</tr>
<tr>
<td>2010</td>
<td>62.124</td>
<td>109.133</td>
<td>9</td>
<td>5.21e-10</td>
</tr>
<tr>
<td>2011</td>
<td>78.398</td>
<td>118.691</td>
<td>9</td>
<td>3.36e-13</td>
</tr>
<tr>
<td>2012</td>
<td>147.176</td>
<td>172.373</td>
<td>9</td>
<td>2.20e-19</td>
</tr>
<tr>
<td>2013</td>
<td>58.429</td>
<td>163.995</td>
<td>9</td>
<td>2.69e-09</td>
</tr>
<tr>
<td>2014</td>
<td>62.136</td>
<td>172.312</td>
<td>9</td>
<td>5.19e-09</td>
</tr>
</tbody>
</table>

Alternative hypothesis: variances are not identical

Table 4.7: Brown Forsyth Test of Homogeneity of variance

<table>
<thead>
<tr>
<th>YEAR</th>
<th>F-Statistic</th>
<th>DF</th>
<th>Region</th>
<th>DF</th>
<th>Residuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maize</td>
<td>Rice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td>4.0249</td>
<td>8.1244</td>
<td>9</td>
<td>136</td>
<td>1.38e-04</td>
</tr>
<tr>
<td>2006</td>
<td>4.9686</td>
<td>7.051</td>
<td>9</td>
<td>128</td>
<td>9.89e-06</td>
</tr>
<tr>
<td>2007</td>
<td>2.7202</td>
<td>5.1463</td>
<td>9</td>
<td>128</td>
<td>6.19e-03</td>
</tr>
<tr>
<td>2008</td>
<td>5.5392</td>
<td>8.6752</td>
<td>9</td>
<td>128</td>
<td>1.97e-06</td>
</tr>
<tr>
<td>2010</td>
<td>4.939</td>
<td>11.4745</td>
<td>9</td>
<td>128</td>
<td>1.08e-05</td>
</tr>
<tr>
<td>2012</td>
<td>0.5799</td>
<td>9.0994</td>
<td>9</td>
<td>158</td>
<td>0.8122</td>
</tr>
<tr>
<td>2013</td>
<td>2.7309</td>
<td>8.1734</td>
<td>9</td>
<td>160</td>
<td>0.00545</td>
</tr>
</tbody>
</table>

Alternative hypothesis: variances are not identical

Table 4.8a: Intercorrelation of maize yields

<table>
<thead>
<tr>
<th>(Intr)</th>
<th>BRONG AHAFO</th>
<th>CENTRAL</th>
<th>EASTERN</th>
<th>GREATER ACCRA</th>
<th>NORTHERN</th>
<th>UPPER EAST</th>
<th>UPPER WEST</th>
<th>VOLTA</th>
<th>WESTERN</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRONG AHAFO</td>
<td>-0.699</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CENTRAL</td>
<td>-0.595</td>
<td>0.416</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EASTERN</td>
<td>-0.673</td>
<td>0.470</td>
<td>0.401</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GREATER ACCRA</td>
<td>-0.546</td>
<td>0.381</td>
<td>0.325</td>
<td>0.368</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NORTHERN</td>
<td>-0.698</td>
<td>0.488</td>
<td>0.415</td>
<td>0.470</td>
<td>0.381</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UPPER EAST</td>
<td>-0.535</td>
<td>0.374</td>
<td>0.318</td>
<td>0.360</td>
<td>0.292</td>
<td>0.373</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UPPER WEST</td>
<td>-0.545</td>
<td>0.381</td>
<td>0.324</td>
<td>0.367</td>
<td>0.298</td>
<td>0.381</td>
<td>0.292</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VOLTA</td>
<td>-0.642</td>
<td>0.449</td>
<td>0.382</td>
<td>0.432</td>
<td>0.351</td>
<td>0.448</td>
<td>0.344</td>
<td>0.384</td>
<td></td>
</tr>
<tr>
<td>WESTERN</td>
<td>-0.598</td>
<td>0.418</td>
<td>0.356</td>
<td>0.403</td>
<td>0.327</td>
<td>0.418</td>
<td>0.320</td>
<td>0.642</td>
<td>0.598</td>
</tr>
</tbody>
</table>
## Table 4.9: Pillai's Test

<table>
<thead>
<tr>
<th>Year</th>
<th>Estimates</th>
<th>DF</th>
<th>Approx.F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>1.0741</td>
<td>9</td>
<td>13.148</td>
<td>2.201e-16***</td>
</tr>
<tr>
<td>2006</td>
<td>1.1379</td>
<td>9</td>
<td>18.772</td>
<td>2.211e-16***</td>
</tr>
<tr>
<td>2007</td>
<td>0.9189</td>
<td>9</td>
<td>12.088</td>
<td>2.234e-16***</td>
</tr>
<tr>
<td>2008</td>
<td>1.0420</td>
<td>9</td>
<td>11.845</td>
<td>2.231e-16***</td>
</tr>
<tr>
<td>2009</td>
<td>0.9501</td>
<td>9</td>
<td>12.870</td>
<td>2.220e-16***</td>
</tr>
<tr>
<td>2010</td>
<td>0.9069</td>
<td>9</td>
<td>11.800</td>
<td>2.242e-16***</td>
</tr>
<tr>
<td>2011</td>
<td>1.1942</td>
<td>9</td>
<td>21.075</td>
<td>2.247e-16***</td>
</tr>
<tr>
<td>2012</td>
<td>0.5287</td>
<td>9</td>
<td>6.3077</td>
<td>2.874e-13***</td>
</tr>
<tr>
<td>2013</td>
<td>0.8559</td>
<td>9</td>
<td>13.302</td>
<td>2.210e-16***</td>
</tr>
</tbody>
</table>
## Table 4.10: Hotelling-Lawley Trace

<table>
<thead>
<tr>
<th>Year</th>
<th>Estimates</th>
<th>DF</th>
<th>Approx.F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>4.8065</td>
<td>9</td>
<td>26.703</td>
<td>2.201e-16***</td>
</tr>
<tr>
<td>2006</td>
<td>2.9192</td>
<td>9</td>
<td>20.435</td>
<td>2.213e-16***</td>
</tr>
<tr>
<td>2007</td>
<td>2.9036</td>
<td>9</td>
<td>20.325</td>
<td>2.231e-16***</td>
</tr>
<tr>
<td>2008</td>
<td>2.1833</td>
<td>9</td>
<td>11.645</td>
<td>2.212e-16***</td>
</tr>
<tr>
<td>2009</td>
<td>1.8931</td>
<td>9</td>
<td>13.252</td>
<td>2.223e-16***</td>
</tr>
<tr>
<td>2010</td>
<td>1.7795</td>
<td>9</td>
<td>12.457</td>
<td>2.202e-16***</td>
</tr>
<tr>
<td>2011</td>
<td>3.1524</td>
<td>9</td>
<td>22.067</td>
<td>2.240e-16***</td>
</tr>
<tr>
<td>2012</td>
<td>0.5365</td>
<td>9</td>
<td>6.3713</td>
<td>2.062e-13***</td>
</tr>
<tr>
<td>2013</td>
<td>1.5253</td>
<td>9</td>
<td>13.389</td>
<td>2.239e-16***</td>
</tr>
<tr>
<td>2014</td>
<td>1.6510</td>
<td>9</td>
<td>14.492</td>
<td>2.200e-16***</td>
</tr>
</tbody>
</table>

## Table 4.11: Roy’s Greatest Root Test

<table>
<thead>
<tr>
<th>Year</th>
<th>Estimates</th>
<th>DF</th>
<th>Approx.F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>4.4601</td>
<td>9</td>
<td>50.548</td>
<td>2.231e-16***</td>
</tr>
<tr>
<td>2006</td>
<td>2.0458</td>
<td>9</td>
<td>29.096</td>
<td>2.233e-16***</td>
</tr>
<tr>
<td>2007</td>
<td>2.6666</td>
<td>9</td>
<td>37.924</td>
<td>2.221e-16***</td>
</tr>
<tr>
<td>2008</td>
<td>1.1822</td>
<td>9</td>
<td>12.873</td>
<td>2.548e-13***</td>
</tr>
<tr>
<td>2009</td>
<td>1.2314</td>
<td>9</td>
<td>17.513</td>
<td>2.211e-16***</td>
</tr>
<tr>
<td>2010</td>
<td>1.2267</td>
<td>9</td>
<td>17.446</td>
<td>2.213e-16***</td>
</tr>
<tr>
<td>2011</td>
<td>2.0693</td>
<td>9</td>
<td>29.430</td>
<td>1.224e-16***</td>
</tr>
<tr>
<td>2012</td>
<td>0.4989</td>
<td>9</td>
<td>8.7591</td>
<td>1.254e-10***</td>
</tr>
<tr>
<td>2013</td>
<td>0.9222</td>
<td>9</td>
<td>16.394</td>
<td>2.221e-16***</td>
</tr>
<tr>
<td>2014</td>
<td>0.9144</td>
<td>9</td>
<td>16.257</td>
<td>2.244e-16***</td>
</tr>
</tbody>
</table>

## Table 4.12: Wilks’ Lambda Test

<table>
<thead>
<tr>
<th>Year</th>
<th>Estimates</th>
<th>DF</th>
<th>Approx.F</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>0.1360</td>
<td>9</td>
<td>19.205</td>
<td>2.212e-16***</td>
</tr>
<tr>
<td>2006</td>
<td>0.1753</td>
<td>9</td>
<td>19.597</td>
<td>2.220e-16***</td>
</tr>
<tr>
<td>2007</td>
<td>0.2205</td>
<td>9</td>
<td>15.942</td>
<td>2.214e-16***</td>
</tr>
<tr>
<td>2008</td>
<td>0.2290</td>
<td>9</td>
<td>11.745</td>
<td>2.231e-16***</td>
</tr>
<tr>
<td>2009</td>
<td>0.2697</td>
<td>9</td>
<td>13.061</td>
<td>2.213e-16***</td>
</tr>
<tr>
<td>2010</td>
<td>0.2892</td>
<td>9</td>
<td>12.128</td>
<td>2.210e-16***</td>
</tr>
<tr>
<td>2011</td>
<td>0.1564</td>
<td>9</td>
<td>21.570</td>
<td>2.226e-16***</td>
</tr>
<tr>
<td>2012</td>
<td>0.7424</td>
<td>9</td>
<td>6.4341</td>
<td>1.490e-13***</td>
</tr>
<tr>
<td>2013</td>
<td>0.3245</td>
<td>9</td>
<td>13.389</td>
<td>2.221e-16***</td>
</tr>
<tr>
<td>2014</td>
<td>0.3008</td>
<td>9</td>
<td>14.545</td>
<td>2.237e-16***</td>
</tr>
</tbody>
</table>
### Appendix C: Approximate 95% confidence intervals

#### Table 4.13a 95% confidence intervals for maize yields

<table>
<thead>
<tr>
<th>Fixed effects:</th>
<th>lower</th>
<th>est.</th>
<th>upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.233719566</td>
<td>1.371765247</td>
<td>1.50981093</td>
</tr>
<tr>
<td>REGIONBRONG AHAFO</td>
<td>0.510918851</td>
<td>0.711442215</td>
<td>0.91196558</td>
</tr>
<tr>
<td>REGIONCENTRAL</td>
<td>0.263163392</td>
<td>0.490303632</td>
<td>0.71744387</td>
</tr>
<tr>
<td>REGIONEASTERN</td>
<td>0.107238693</td>
<td>0.306808230</td>
<td>0.50637777</td>
</tr>
<tr>
<td>REGIONGREATER ACCRA</td>
<td>-0.768529813</td>
<td>-0.51681752</td>
<td>-0.26423369</td>
</tr>
<tr>
<td>REGIONNORTHERN</td>
<td>-0.181427607</td>
<td>0.016931401</td>
<td>0.21529041</td>
</tr>
<tr>
<td>REGIONUPPER EAST</td>
<td>-0.596581724</td>
<td>-0.362592412</td>
<td>-0.12860310</td>
</tr>
<tr>
<td>REGIONUPPER WEST</td>
<td>-0.404258743</td>
<td>-0.148988117</td>
<td>0.10628251</td>
</tr>
<tr>
<td>REGIONVOLTA</td>
<td>-0.156168130</td>
<td>0.060506982</td>
<td>0.27718209</td>
</tr>
<tr>
<td>REGIONWESTERN</td>
<td>-0.197297246</td>
<td>0.026169402</td>
<td>0.24963605</td>
</tr>
<tr>
<td>ASHANTI:YEAR</td>
<td>-0.020172801</td>
<td>-0.004539020</td>
<td>0.01109476</td>
</tr>
<tr>
<td>REGIONBRONG AHAFO:YEAR</td>
<td>-0.033025601</td>
<td>-0.010648155</td>
<td>0.01172929</td>
</tr>
<tr>
<td>REGIONCENTRAL:YEAR</td>
<td>-0.019481393</td>
<td>0.006795677</td>
<td>0.03307275</td>
</tr>
<tr>
<td>REGIONEASTERN:YEAR</td>
<td>0.039720357</td>
<td>0.062942233</td>
<td>0.08616411</td>
</tr>
<tr>
<td>REGIONGREATER ACCRA:YEAR</td>
<td>0.019455597</td>
<td>0.048127372</td>
<td>0.07679915</td>
</tr>
<tr>
<td>REGIONNORTHERN:YEAR</td>
<td>-0.009439387</td>
<td>0.012956791</td>
<td>0.03533297</td>
</tr>
<tr>
<td>REGIONUPPER EAST:YEAR</td>
<td>0.022047189</td>
<td>0.051288249</td>
<td>0.08052931</td>
</tr>
<tr>
<td>REGIONUPPER WEST:YEAR</td>
<td>0.012070049</td>
<td>0.040755118</td>
<td>0.06940199</td>
</tr>
<tr>
<td>REGIONVOLTA:YEAR</td>
<td>0.011513308</td>
<td>0.035855783</td>
<td>0.06019826</td>
</tr>
<tr>
<td>REGIONWESTERN:YEAR</td>
<td>-0.030273537</td>
<td>-0.004150291</td>
<td>0.02197295</td>
</tr>
</tbody>
</table>

**Random Effects:**
- **Level:** DISTRICT
  - lower est. upper
  - sd((Intercept)) 0.2179021 0.2572860 0.2850493

**Within-group standard error:**
- lower est. upper
  - 0.309047 0.3223060 0.3353117

#### Table 4.13b 95% confidence intervals for rice yields

<table>
<thead>
<tr>
<th>Fixed effects:</th>
<th>lower</th>
<th>est.</th>
<th>upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.75027625</td>
<td>1.03944940</td>
<td>1.31961294</td>
</tr>
<tr>
<td>REGIONBRONG AHAFO</td>
<td>0.06886600</td>
<td>0.48628973</td>
<td>0.90371347</td>
</tr>
<tr>
<td>REGIONCENTRAL</td>
<td>-0.1436250</td>
<td>0.32029622</td>
<td>0.78395494</td>
</tr>
<tr>
<td>REGIONEASTERN</td>
<td>0.17289168</td>
<td>0.58338800</td>
<td>0.99388431</td>
</tr>
<tr>
<td>REGIONGREATER ACCRA</td>
<td>0.36426868</td>
<td>0.87728195</td>
<td>1.39029522</td>
</tr>
<tr>
<td>REGIONNORTHERN</td>
<td>0.61942247</td>
<td>1.03272261</td>
<td>1.44602275</td>
</tr>
<tr>
<td>REGIONUPPER EAST</td>
<td>1.10541142</td>
<td>1.58096863</td>
<td>2.05652584</td>
</tr>
<tr>
<td>REGIONUPPER WEST</td>
<td>-0.37704547</td>
<td>0.15518159</td>
<td>0.68740865</td>
</tr>
<tr>
<td>REGIONVOLTA</td>
<td>0.26529335</td>
<td>0.71409241</td>
<td>1.16289148</td>
</tr>
<tr>
<td>REGIONWESTERN</td>
<td>-0.25088510</td>
<td>0.20516262</td>
<td>0.66121034</td>
</tr>
<tr>
<td>ASHANTI:YEAR</td>
<td>0.15890483</td>
<td>0.18677354</td>
<td>0.21464226</td>
</tr>
<tr>
<td>REGIONBRONG AHAFO:YEAR</td>
<td>-0.19361817</td>
<td>-0.15389694</td>
<td>-0.11417570</td>
</tr>
<tr>
<td>REGIONCENTRAL:YEAR</td>
<td>-0.17573418</td>
<td>-0.12846772</td>
<td>-0.08120126</td>
</tr>
</tbody>
</table>
Random Effects:
Level: DISTRICT

lower est. upper
sd((Intercept)) 0.5419314 0.6093177 0.685083

within-group standard error:
lower est. upper

0.5431437 0.5651979 0.5881477

Appendix D: Marginal Variance Covariance Matrix

TABLE 4.14: MAIZE: MARGINAL VARIANCE COVARIANCE MATRIX

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.16669</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
</tr>
<tr>
<td>2</td>
<td>0.062832</td>
<td>0.16669</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
</tr>
<tr>
<td>3</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.16669</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
</tr>
<tr>
<td>4</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.16669</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
</tr>
<tr>
<td>5</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.16669</td>
<td>0.062832</td>
<td>0.062832</td>
</tr>
<tr>
<td>6</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.16669</td>
<td>0.062832</td>
</tr>
<tr>
<td>7</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.062832</td>
<td>0.16669</td>
</tr>
</tbody>
</table>

Random Effects Variance Covariance Matrix
(Intercept)
(Intercept) 0.062832

Standard Deviations: 0.25066

TABLE 4.15: RICE: MARGINAL VARIANCE COVARIANCE MATRIX

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6904</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
</tr>
<tr>
<td>2</td>
<td>0.37098</td>
<td>0.6904</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
</tr>
<tr>
<td>3</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.6904</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
</tr>
<tr>
<td>4</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.6904</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
</tr>
<tr>
<td>5</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.6904</td>
<td>0.37098</td>
<td>0.37098</td>
</tr>
<tr>
<td>6</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.6904</td>
<td>0.37098</td>
</tr>
<tr>
<td>7</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.37098</td>
<td>0.6904</td>
</tr>
</tbody>
</table>
Random Effects Variance Covariance Matrix
(Intercept)
(Intercept) 0.37098
Standard Deviations: 0.60908

Appendix E: Pairwise Comparisons of difference in Mean Yields

Table 4.16a: Pairwise Comparisons in difference of mean yields

<table>
<thead>
<tr>
<th>(I) Region</th>
<th>(J) Region</th>
<th>Differences in Mean yields (I-J)</th>
<th>Std. Error</th>
<th>P-value</th>
<th>95% Confidence Interval for Difference</th>
<th>95% Confidence Interval for Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Maize</td>
<td>Rice</td>
<td>Maize</td>
<td>Rice</td>
<td>Maize</td>
</tr>
<tr>
<td>ASHANTI</td>
<td>B/AHAFO</td>
<td>-.613*</td>
<td>-.449*</td>
<td>.039</td>
<td>.074</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>CENTRAL</td>
<td>-.548*</td>
<td>-.565*</td>
<td>.043</td>
<td>.082</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>EASTERN</td>
<td>-.701*</td>
<td>-.350*</td>
<td>.040</td>
<td>.076</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>G/ACCRA</td>
<td>.280*</td>
<td>-1.400*</td>
<td>.053</td>
<td>.101</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>NORTHERN</td>
<td>-.130*</td>
<td>0.090</td>
<td>.040</td>
<td>.075</td>
<td>.049</td>
</tr>
<tr>
<td></td>
<td>UPPER EAST</td>
<td>-.009</td>
<td>-.469*</td>
<td>.052</td>
<td>.098</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>UPPER WEST</td>
<td>-.130</td>
<td>.684*</td>
<td>.052</td>
<td>.098</td>
<td>.553</td>
</tr>
<tr>
<td></td>
<td>VOLTA</td>
<td>-.287*</td>
<td>-.620*</td>
<td>.042</td>
<td>.079</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>WESTERN</td>
<td>-.035</td>
<td>.902*</td>
<td>.043</td>
<td>.081</td>
<td>1.000</td>
</tr>
<tr>
<td>B/AHAFO</td>
<td>ASHANTI</td>
<td>.613*</td>
<td>-.449*</td>
<td>.039</td>
<td>.074</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>CENTRAL</td>
<td>.065</td>
<td>.115</td>
<td>.044</td>
<td>.084</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>EASTERN</td>
<td>-.088</td>
<td>-.800*</td>
<td>.041</td>
<td>.078</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>G/ACCRA</td>
<td>.893*</td>
<td>-.1850*</td>
<td>.054</td>
<td>.103</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>NORTHERN</td>
<td>.483*</td>
<td>-.359*</td>
<td>.041</td>
<td>.077</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>UPPER EAST</td>
<td>.604*</td>
<td>-.919*</td>
<td>.053</td>
<td>.099</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>UPPER WEST</td>
<td>.483*</td>
<td>.235</td>
<td>.053</td>
<td>.100</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>VOLTA</td>
<td>.326*</td>
<td>-.1069*</td>
<td>.043</td>
<td>.081</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>WESTERN</td>
<td>.578*</td>
<td>.452*</td>
<td>.044</td>
<td>.084</td>
<td>1.000</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>ASHANTI</td>
<td>.548*</td>
<td>-.565*</td>
<td>.043</td>
<td>.082</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>B/AHAFO</td>
<td>-.065</td>
<td>-.115</td>
<td>.044</td>
<td>.084</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>EASTERN</td>
<td>-.153*</td>
<td>-.915*</td>
<td>.045</td>
<td>.085</td>
<td>.034</td>
</tr>
<tr>
<td></td>
<td>G/ACCRA</td>
<td>.827*</td>
<td>-1.965*</td>
<td>.057</td>
<td>.108</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>NORTHERN</td>
<td>.418*</td>
<td>-.475*</td>
<td>.045</td>
<td>.085</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>UPPER EAST</td>
<td>.538*</td>
<td>-.1034*</td>
<td>.056</td>
<td>.105</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>UPPER WEST</td>
<td>.418*</td>
<td>.11919</td>
<td>.056</td>
<td>.106</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>VOLTA</td>
<td>-.260*</td>
<td>-.1185*</td>
<td>.047</td>
<td>.088</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>WESTERN</td>
<td>.513*</td>
<td>.337*</td>
<td>.048</td>
<td>.091</td>
<td>.000</td>
</tr>
<tr>
<td>EASTERN</td>
<td>ASHANTI</td>
<td>.701*</td>
<td>.350*</td>
<td>.040</td>
<td>.076</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>B/AHAFO</td>
<td>.088</td>
<td>.800*</td>
<td>.041</td>
<td>.078</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>CENTRAL</td>
<td>.153*</td>
<td>.915*</td>
<td>.045</td>
<td>.085</td>
<td>.034</td>
</tr>
<tr>
<td>Region</td>
<td>Intercept</td>
<td>B/AHAFO</td>
<td>NORTHERN</td>
<td>UPPER EAST</td>
<td>UPPER WEST</td>
<td>VOLTA</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------</td>
<td>----------</td>
<td>----------</td>
<td>------------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>G/ ACCRA</td>
<td>-.280</td>
<td>1.400**</td>
<td>.053</td>
<td>.000</td>
<td>.000</td>
<td>.056</td>
</tr>
<tr>
<td>B/ AHAFO</td>
<td>-.893</td>
<td>.185**</td>
<td>.054</td>
<td>.000</td>
<td>.000</td>
<td>.064</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>-.827</td>
<td>1.965*</td>
<td>.057</td>
<td>.000</td>
<td>.000</td>
<td>.064</td>
</tr>
<tr>
<td>EASTERN</td>
<td>-.980</td>
<td>1.050*</td>
<td>.055</td>
<td>.000</td>
<td>.000</td>
<td>.055</td>
</tr>
<tr>
<td>NORTHERN</td>
<td>-.409</td>
<td>1.490*</td>
<td>.055</td>
<td>.000</td>
<td>.000</td>
<td>.055</td>
</tr>
<tr>
<td>UPPER EAST</td>
<td>-.289</td>
<td>.931*</td>
<td>.064</td>
<td>.000</td>
<td>.000</td>
<td>.064</td>
</tr>
<tr>
<td>UPPER WEST</td>
<td>-.410</td>
<td>2.084*</td>
<td>.064</td>
<td>.000</td>
<td>.000</td>
<td>.064</td>
</tr>
<tr>
<td>VOLTA</td>
<td>-.567*</td>
<td>-.780*</td>
<td>.056</td>
<td>.000</td>
<td>.000</td>
<td>.056</td>
</tr>
<tr>
<td>WESTERN</td>
<td>-.314*</td>
<td>2.302*</td>
<td>.057</td>
<td>.000</td>
<td>.000</td>
<td>.057</td>
</tr>
<tr>
<td>NORTHERN</td>
<td>.130</td>
<td>-.0899</td>
<td>.040</td>
<td>.000</td>
<td>.000</td>
<td>.040</td>
</tr>
<tr>
<td>B/ AHAFO</td>
<td>-.483*</td>
<td>.359*</td>
<td>.041</td>
<td>.000</td>
<td>.000</td>
<td>.041</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>-.418*</td>
<td>.475*</td>
<td>.045</td>
<td>.000</td>
<td>.000</td>
<td>.045</td>
</tr>
<tr>
<td>EASTERN</td>
<td>-.571*</td>
<td>-.440*</td>
<td>.042</td>
<td>.000</td>
<td>.000</td>
<td>.042</td>
</tr>
<tr>
<td>G/ ACCRA</td>
<td>.409*</td>
<td>1.490*</td>
<td>.055</td>
<td>.000</td>
<td>.000</td>
<td>.055</td>
</tr>
<tr>
<td>UPPER EAST</td>
<td>.121</td>
<td>-.559*</td>
<td>.053</td>
<td>.000</td>
<td>.000</td>
<td>.053</td>
</tr>
<tr>
<td>UPPER WEST</td>
<td>.000</td>
<td>.594*</td>
<td>.053</td>
<td>.000</td>
<td>.000</td>
<td>.053</td>
</tr>
<tr>
<td>VOLTA</td>
<td>-.157*</td>
<td>-.710*</td>
<td>.043</td>
<td>.000</td>
<td>.000</td>
<td>.043</td>
</tr>
<tr>
<td>WESTERN</td>
<td>.095</td>
<td>.812*</td>
<td>.045</td>
<td>.000</td>
<td>.000</td>
<td>.045</td>
</tr>
<tr>
<td>U/EAST</td>
<td>.009</td>
<td>.469*</td>
<td>.052</td>
<td>.000</td>
<td>.000</td>
<td>.052</td>
</tr>
<tr>
<td>B/ AHAFO</td>
<td>-.604*</td>
<td>.919*</td>
<td>.053</td>
<td>.000</td>
<td>.000</td>
<td>.053</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>-.538*</td>
<td>1.034*</td>
<td>.056</td>
<td>.000</td>
<td>.000</td>
<td>.056</td>
</tr>
<tr>
<td>EASTERN</td>
<td>-.691*</td>
<td>.119*</td>
<td>.053</td>
<td>.000</td>
<td>.000</td>
<td>.053</td>
</tr>
<tr>
<td>G/ ACCRA</td>
<td>.289*</td>
<td>-.931*</td>
<td>.064</td>
<td>.000</td>
<td>.000</td>
<td>.064</td>
</tr>
<tr>
<td>NORTHERN</td>
<td>-.121</td>
<td>1.559*</td>
<td>.053</td>
<td>.000</td>
<td>.000</td>
<td>.053</td>
</tr>
<tr>
<td>UPPER EAST</td>
<td>-.121</td>
<td>1.153*</td>
<td>.063</td>
<td>.000</td>
<td>.000</td>
<td>.063</td>
</tr>
<tr>
<td>UPPER WEST</td>
<td>-.278*</td>
<td>.151*</td>
<td>.055</td>
<td>.000</td>
<td>.000</td>
<td>.055</td>
</tr>
<tr>
<td>VOLTA</td>
<td>-.025*</td>
<td>1.371*</td>
<td>.056</td>
<td>.000</td>
<td>.000</td>
<td>.056</td>
</tr>
<tr>
<td>WESTERN</td>
<td>.130</td>
<td>-.684*</td>
<td>.052</td>
<td>.000</td>
<td>.000</td>
<td>.052</td>
</tr>
<tr>
<td>U/WEST</td>
<td>.130</td>
<td>-.684*</td>
<td>.052</td>
<td>.000</td>
<td>.000</td>
<td>.052</td>
</tr>
<tr>
<td>B/ AHAFO</td>
<td>-.483*</td>
<td>-.234*</td>
<td>.053</td>
<td>.000</td>
<td>.000</td>
<td>.053</td>
</tr>
<tr>
<td>CENTRAL</td>
<td>-.418*</td>
<td>-.1192</td>
<td>.056</td>
<td>.000</td>
<td>.000</td>
<td>.056</td>
</tr>
<tr>
<td>EASTERN</td>
<td>-.571*</td>
<td>-.1034*</td>
<td>.054</td>
<td>.000</td>
<td>.000</td>
<td>.054</td>
</tr>
<tr>
<td>G/ ACCRA</td>
<td>.410*</td>
<td>-2.084*</td>
<td>.064</td>
<td>.000</td>
<td>.000</td>
<td>.064</td>
</tr>
<tr>
<td>NORTHERN</td>
<td>.000</td>
<td>-.594*</td>
<td>.053</td>
<td>.000</td>
<td>.000</td>
<td>.053</td>
</tr>
<tr>
<td>UPPER EAST</td>
<td>.121</td>
<td>-.1153*</td>
<td>.063</td>
<td>.000</td>
<td>.000</td>
<td>.063</td>
</tr>
<tr>
<td>VOLTA</td>
<td>-.157</td>
<td>-1.304*</td>
<td>.055</td>
<td>.000</td>
<td>.000</td>
<td>.055</td>
</tr>
<tr>
<td>WESTERN</td>
<td>0.095</td>
<td>.218*</td>
<td>.056</td>
<td>.000</td>
<td>.000</td>
<td>.056</td>
</tr>
<tr>
<td>VOLTA</td>
<td>0.287</td>
<td>0.620*</td>
<td>.042</td>
<td>.000</td>
<td>.000</td>
<td>.042</td>
</tr>
<tr>
<td>B/ AHAFO</td>
<td>-0.326*</td>
<td>1.069*</td>
<td>.043</td>
<td>.000</td>
<td>.000</td>
<td>.043</td>
</tr>
</tbody>
</table>

** denotes statistical significance at the 0.05 level.
<table>
<thead>
<tr>
<th>REGION</th>
<th>ASHANTI</th>
<th>B/A</th>
<th>CENTRAL</th>
<th>E/R</th>
<th>G/A</th>
<th>N/R</th>
<th>U/EAST</th>
<th>U/WEST</th>
<th>VOLTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>B/AHAFO</td>
<td>&lt;2e-16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CENTRAL</td>
<td>&lt;2e-16</td>
<td>1.00000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EASTERN</td>
<td>&lt;2e-16</td>
<td>1.00000</td>
<td>0.06824</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G/ACCRA</td>
<td>1.10e-05</td>
<td>&lt;2e-16</td>
<td>&lt;2e-16</td>
<td>&lt;2e-16</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NORTHERN</td>
<td>0.067220</td>
<td>&lt;2e-16</td>
<td>&lt;2e-16</td>
<td>&lt;2e-16</td>
<td>1.30e-11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U/EAST</td>
<td>1.000000</td>
<td>&lt;2e-16</td>
<td>&lt;2e-16</td>
<td>&lt;2e-16</td>
<td>0.000480</td>
<td>1.00000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U/WEST</td>
<td>0.716450</td>
<td>&lt;2e-16</td>
<td>&lt;2e-16</td>
<td>&lt;2e-16</td>
<td>2.00e-08</td>
<td>1.00000</td>
<td>1.00000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VOLTA</td>
<td>5.00e-10</td>
<td>3.60e-12</td>
<td>4.40e-07</td>
<td>&lt;2e-16</td>
<td>&lt;2e-16</td>
<td>1.20e-02</td>
<td>1.90e-05</td>
<td>0.15084</td>
<td></td>
</tr>
<tr>
<td>WESTERN</td>
<td>1.000000</td>
<td>&lt;2e-16</td>
<td>&lt;2e-16</td>
<td>&lt;2e-16</td>
<td>2.60e-06</td>
<td>1.00000</td>
<td>1.000000</td>
<td>1.00000</td>
<td>4.10e-06</td>
</tr>
</tbody>
</table>

Source: SRID of MOFA
Appendix F: R Codes used for the study

R CODES FOR MANOVA MODEL

data1<-read.csv("mixed2.csv",header=TRUE)

plot(data1)

##MANOVA MODEL

crop_manova <- manova(cbind(MAIZE, RICE) ~ REGION, data = data1)

summary(crop_manova)

##EXPLORATORY ANALYSIS, TEST OF ASSUMPTIONS

RESID<-resid(crop_manova)

##BOXPLOT

boxplot(RESID,main="Boxplot of Residuals",ylim=c(-2,3),col="lightgreen")

##linearity assumption

maize<-data1$MAIZE
rice<-data1$RICE

par(mfrow=c(1,2))

plot(maize, main=" Scatterplot of maize")

plot(rice,main="Scatterplot of rice")

library(MVN)

par(mfrow=c(2,2))

uniPlot(RESID,type="qqplot")

uniPlot(RESID,type = "histogram",)

##Homogeneity of Variance Plot

bartlett.test(MAIZE~REGION, data=data1)

bartlett.test(RICE~REGION, data=data1)
library(HH)

hov(MAIZE~REGION, data=data1)
hov(RICE~REGION, data=data1)
par(mfrow=c(2,2))
hovPlot(MAIZE~REGION, data=data1)
hovPlot(RICE~REGION, data=data1)

## Test of differences in crop yields using manova
summary(crop_manova, test = "Pillai")
summary(crop_manova, test = "Wilks")
summary(crop_manova, test = "Hotelling-Lawley")
summary(crop_manova, test = "Roy")
summary.aov(crop_manova)

## We might now move on to investigate the difference in yields across the regions

## Bonferroni post hoc for maize MAIZE
attach(data1)
group<-.as.factor(REGION)
MM<-tapply(maize, REGION, mean)
diff<-pairwise.t.test(MAIZE, REGION, p.adjust="bonferroni")
diff

## Bonferroni post hoc for rice MAIZE

## Tukey HSD Post hoc test for maize yields
par(mfrow=c(1,1))
library(TukeyC)
manova_aov<-aov(MAIZE~REGION, data=data1)
CI<-TukeyHSD(manova_aov, "REGION")
CI
plot(CI)

## Tukey HSD Post hoc test for rice yields
man_aov<-aov(RICE~REGION,data=data1)
CONFI<-TukeyHSD(man_aov, "REGION")
CONFI
plot(CONFI)

## Box M Test
library(biotools)
head(data1)
box M (data1[,4], data1[,5])

R CODES FOR LINEAR MIXED MODEL

data<-read.csv("mixed.csv",header=TRUE)
dm<-str(data)
head(data)

# Exploratory Analysis
summary(data)
maize<-data$MAIZE
rice<-data$RICE
par(mfrow=c(2,2))
boxplot(maize, main="Boxplot for Maize yield",ylab = "Maize yield in Mt/Ha", col="orange")
qqnorm(maize, main = "Normal Q-Q Plot for Maize yields", ylab="Maize yield in Mt/Ha")
qqline(maize)

boxplot(rice, main= "Boxplot for Rice yield", ylab = "Rice yield in Mt/Ha",col="lightgreen")
qqnorm(rice, main = "Normal Q-Q Plot for Rice yields", ylab="Rice yield in Mt/Ha")
qqline(maize)
REGION<--data$REGION
par(mfrow=c(2,2))
boxplot(MAIZE ~ REGION, data=data,col="orange", xlab="REGION", ylab="Maize (Mt/Ha)", ylim=c(0,6),main="Maize yields by Region")
boxplot(RICE ~ REGION, data=data,col="orange", xlab="REGION", ylab="Rice (Mt/Ha)", ylim=c(0,8), main="Rice yields by Region")
boxplot(MAIZE ~ YEAR, data=data,col="orange", xlab="YEAR", ylab="Maize (Mt/Ha)", ylim=c(0,6),main="Maize yields by Year")
boxplot(RICE ~ YEAR, data=data,col="orange", xlab="YEAR", ylab="Rice (Mt/Ha)", ylim=c(0,6),main="Rice yields by Year")

#LMM MAIZE
## Random intercept
library(lme4)
fit2<-lmer(MAIZE~REGION+YEAR+(1|DISTRICT)+(REGION*YEAR),data)
fit21<-lmer(MAIZE~REGION+YEAR+(1|DISTRICT)+(REGION*YEAR),data,REML=FALSE)
#random intercept and slope
fitlope <-
lmer(MAIZE~REGION+YEAR+(1|YEAR)+(1|DISTRICT)+(REGION*YEAR),data=data,REML=FALSE)
anova(fit21,fitlope)
summary(fitlope)
par(mfrow=c(1,1))
hlmList <- lmList(MAIZE~ YEAR|DISTRICT, data=data)
coefs <- coef(hlmList)
names(coefs) <- c("Intercept", "Slope")
plot(Slope~Intercept, data=coefs)
abline(lm(Slope~Intercept, data=coefs))

#LMM RICE
## Random intercept
library(lme4)
fit3<-lmer(RICE~REGION+YEAR+(1|DISTRICT)+(REGION*YEAR),data)
fit31<-lmer(RICE~REGION+YEAR+(1|DISTRICT)+(REGION*YEAR),data,REML=FALSE)
#random intercept and slope
fitlope2
lmer(RICE~REGION+YEAR+(1|YEAR)+(1|DISTRICT)+(REGION*YEAR),data=data,REML=FALSE)
anova(fit31,fitlope2)
summary(fitlope2)
par(mfrow=c(1,1))
hlmList <- lmList(RICE~ YEAR|DISTRICT, data=data)
coefs <- coef(hlmList)
names(coefs) <- c("Intercept", "Slope")
plot(Slope~Intercept, data=coefs)
abline(lm(Slope~Intercept, data=coefs))
library(car)
par(mfrow=c(2,1))
plot((fit2),main="Scatter plot of Maize yield")
plot((fit3),main="Scatter plot of Rice yield")
residual<-resid(fit2)
residuals<-resid(fit3)
par(mfrow=c(2,2))
hist(residual, ylab="Maize yields(Mt/Ha)",freq=FALSE, main="Histogram of Maize residual", col="orange",ylim=c(0.0,1.0))
curve(dnorm(x, mean=mean(residuals), sd=sd(residuals)),
      add=TRUE, col="darkblue", lwd=2)
hist(residuals,ylab="Rice yields(Mt/Ha)", freq=FALSE,main="Histogram of Rice residual",col="lightgreen", ylim=c(0.0,1.0))
curve(dnorm(x, mean=mean(residual), sd=sd(residual)),
      add=TRUE, col="darkblue", lwd=2)
qqnorm(residual, main="Normal QQ-Plot for Residual of maize");qqline(residual)
qqnorm(residuals,main="Normal QQ-Plot for Residual of Rice");qqline(residuals)