KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

COLLEGE OF SCIENCE

DEPARTMENT OF MATHEMATICS

MULTIPLE COMPARISON AND RANDOM EFFECT MODEL ON COCOA PRODUCTION IN GHANA (FROM 1969/70 TO 2010/11 PRODUCTION YEARS)

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY, KUMASI, IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF MASTER OF PHILOSOPHY DEGREE IN MATHEMATICS

BY

ACHEAMPONG AMO ISAAC

JUNE, 2012.

DECLARATION

I, AcheampongAmo Isaac, author of this thesis do hereby declare that apart from the references of other people's work which has duly been acknowledged, the research work presented in this thesis was done entirely by me at the Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi .

I do hereby declare that, this work has neither been presented in whole nor in part for any degree at this university or elsewhere.

ACHEAMPONG AMO ISAAC	………………………	………………
(STUDENT)	SIGNATURE	DATE
NANA KENA FREMPONG	………………………	...........………..
(SUPERVISOR)	SIGNATURE	DATE
MR. F.K. DARKWAH	……………………	……………….
(HEAD OF DEPARTMENT)	SIGNATURE	DATE

DEDICATION

This thesis is entirely dedicated to Mr. Samuel AppohNyameke, Range Supervisor of Forestry Commission of Ghana in the AtwimaMponua District

ACKNOWLEDGEMENT

If it had not been the Lord who was on my side, now may Isreal say (Psalm 124: 1)it is He that had made me and not myself; may his name be glorified.

I owe an utmost gratitude to my supervisor Nana KenaFrimpong whose useful comments, suggestions and encouragement propelled me to see to the end of this thesis.

I am exceedingly grateful to Dr. OheneBaffour of COCOBOD who offered invaluable contribution, constructive criticisms and suggestions as to how this thesis could be high quality.

I would be very ungrateful if I forget my best friend Mr. OduroKwarteng of KMA for his advice, encouragement and love.

Finally, my sincere gratitude to my wife Mrs. Debora AmoAcheampong whose diverse contribution helped me to shape this work.

ABSTRACT

The study analyses the production of cocoa in Ghana from 1969/70 to 2010/11 .The study area (Ashanti, Western,Brong Ahafo, Central, Eastern and Volta regions) were particularly chosen for this study because of their prime places in cocoa production in Ghana.

The descriptive statistics used to analyze the results showed that majority (56.33%) of cocoa produced in the country comes from western region of Ghana with the mean annual production of 142823.9 metric tonnes while the lowest production(0.5%) of cocoa comes from volta region with mean annual production being 4897.00 metric tonnes.

The ANOVA analysis revealed that there is significant difference(significance= 0.0001) between the mean production of cocoa in the country and further analysis using the multiple comparison (pair wise tests) also showed that the mean annual production of cocoa differ in terms of the six region of cocoa production in the country.

Multiple comparison to detect where the difference lies using the three the comparison tests in the study(Tukey,LSD,Scheffe and Bonfferoni) showed that the mean difference between Ashanti and Western region was not significant but the mean difference between Ashanti and Central, Eastern, Volta ,Western were significant with mean difference of 53679.23,39441.73,83654.48 and -54272.38 respectively.

Further analysis using mixed effect model also revealed that from 1969/70 production, all the six regions with the exception of Volta region experienced increasing cocoa production trend as time increased. Western and Ashanti regions had the highest production over the years.

Random intercept with variance-covariance assumption also showed that the different regions had variations in the mean cocoa production.

TABLE OF CONTENT

CHAPTER ONE

1.1 Background……………………………………………………………………………....1

1.2 Justification……………………………………………………………........................5

1.3 Statement of problem…………………………………………………………………..5

1.4 Research Objectives…………………………………...6

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

1.6 Scope of the study…………………………………………………………………..…7

1.7 Limitation of the study………………………………...8

1.8 Organization of the study…………………………..8

CHAPTER TWO

2.1Origin and Spread of cocoa in Ghana……………...9

2.2 Cocoa Production in Ghana……………………………………………………………...10

2.2.1 The structure of cocoa production………………...……………………….………....10

2.2.2The role of cocoa in Ghana's Economy……………………………...……………..12

2.4 Determinants of cocoa production……………………….…………………...……………14

2.4.1 Control of diseases and pests of cocoa……………………………………..…………..17

2.4.2 Fertilizer Application……………………………….……………..…………..…20

2.4.3 Rainfall……..…………………………………………………………..…………..….21

2.4.4 The Producer Price of Cocoa……………………….……………………………..22

2.4.5 Labour and Other Socio-economic factors………………………………….……....25

2.5 The Concept of Efficiency in Agricultural Production………………………………….25

2.6 Analysis Of Variance……………………………………………………………… .29

2.6.1 Assumptions OfAnova……………………………………………………….............29

2.6.2 Types Of Anova Models……………..………………………………………………...31

2.6.3 Fixed-effects models………………………..…………………………………….....31

2.6.4 Random-effects models…………………………...………………………………………32

2.6.5 Mixed Effect Model………………………………………………………………...........32

2.6.6 Linear Mixed Effects Models……………………………………………….……….33

2.6.7 Nonlinear Mixed Effects Models…………………………………………………..…..34

2.6.8 Multiple Comparison Test (Post Hoc Tests)……………………………………........35

CHAPTER THREE

3.0 Anova (Analysis Of Variance)…………………….……………………………........38

3.1 Factor Effects Model for A Single Factor (Region) Anova…………………………..........40

3.2 Multiple Comparison Test (Post Hoc Tests)……………………………....................42

3.2.1 Experiment – Wise Error Rate……………………….……………….......................42

3.2.2 Tukey's Methods ……………………..…………………………………….…………….44

3.2.3 Fisher's LSD Method…..46

3.2.4 Shceffé Method ForCompairing All Contrasts……………………………….……….48

3.2.4 Bonferroni Method……………………………………………….........................50

3.3 Mixed Effect Model……….…………………………………………….……......52

3.4 Estimation of Parameters in the Mixed Model …………….………….........................54

3.4.1 Maximum Likelihood Estimation………………………………………………......54

3.4.2 Restricted Maximum Likelihood Estimator (Reml) For Mixed Model…....................55

3.5.0 Akaike Information Criterion (Aic) For Model Selection……………………..............57

3.6 Random Effect Models……………………………………………………………...58

CHAPTER FOUR

4.0 Data Presentation And Analysis…………………………….………......................58

4.1 Analysis Of Variance (Anova) On Yield And Year Of Production….……………......60

4.2.1 Assessing Varieties For Assumptions……………………………………..….........61

4.3Transformation……………………………………………………………..…….62

4.4.0 Post Hoc Test………………………………………………………………….…..63

4.5 Contrasts…………………………………………………………………….....64

4.5.1 Establishing Contrast By Geographical Locations…………………………......…..66

4.6 Mixed Effect Model………………………….…………………………….…….…...68

CHAPTER FIVE

5.1Discussion And Conclusion………………..……………………….………………73

5.2 Recommendation…………………………………………….………….……………………75

REFERENCES………………………………………………………………………………76-78

APPENDICES………………………………………………………………………………79-83

.

LIST OF TABLES **PAGES**

TABLE 3.1 ANOVA TABLE FOR ONE WAY ANALYSIS………………………………39

TABLE 4.1 SUMMARY STATISTICS ON YIELD PER REGION ……………………….58

TABLE 4.2 ANOVA TABLE FOR YIELD OF COCOA………………………………...60

TABLE 4.3 ANOVA MODEL FOR YEAR OF PRODUCTION…………………………...60

TABLE 4.4 LEVENE'S TEST FOR HOMOGENEITY OF YIELD VARIANCE………....61

TABLE 4.5 SUMMARY OF THE MULTIPLE COMPARISON TEST……………………64

TABLE 4.6 AIC VALUES FOR DIFFERENT MODELS………………………………..72

LIST OF FIGURES **PAGES**

FIGURE 4.1 NORMAL PROBABILITY PLOT…………………………………………....62

FIGURE 4.2 INDIVIDUAL MEAN OF COCOA PRODUCTION PER REGION………..68

FIGURE 4.3 LOG TRANSFORMATION………………………………………………..80

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Ghana is endowed with a variety of mineral and agricultural product (Breisinger et al, 2008). Historically, Ghana's Gross Domestic Product (GDP), an indicator of the total value of all goods and services produced within Ghana, has been led mostly by the agricultural sector of the economy.

In 2009, agriculture contributed 46.7% of the overall in GDP, up from the 41.4% recorded the previous year. This growth was in large part, a direct consequence of the rehabilitation of cocoa production (Aryeetey and Kanbur, 2008).

Africa's first crop of cocoa was planted in Ghana over a century ago by Tetteh Quarshie (COCOBOD), 2000). Subsequently, Ghana's status as a cocoa producer grew and peaked in the mid 1960s, collapsed in the early 1780s, and was revived in the late 1980s, (Jaeger, 1999). In the 2002/2003 season, the country's cocoa production peaked similar to the 1960 levels. Ghana's output reached 566,000 tonnes in the mid 1960s before falling to about 159,000 tonnes I the early 1980s. Ghana's cocoa output increased to 350,000 tonnes at the end of 1999. Due to good agronomic practices and higher cocoa prices, the output reached 700,000 tonnes in 2008(COCOBOD 2009).

The fall in cocoa output in Ghana's history was attributed mainly to poor management of the Cocoa Marketing Board (Jaeger, 1999). However, several other reasons can also be cited. The main causes include land degradation in the producing area and swollen shoot disease (Jaeger, 1999). A fall in international terms of trade in the late 1970s and early 1980s resulting in the decline in cocoa price also affected production.

(Jaeger, 1999). Another cause of the fall was the drought and extensive bush fire in the 1980s, the massive migration to Nigeria due to poor economic condition domestically and the oil boom in Nigeria .

Consequently, output per worker plummeted compared to historical levels, reaching 4.17% between 1980 and 1984 (Aryttey and Kanbur, 2008). Since cocoa production is a major contributor to the economy, it was only natural for the GDP of the country to fall as production declined (Argeetey and Kanbur, 2008).

Analysis of the trends of Ghana's history reveals that increased cocoa revenue is associated with rising economic growth (Brempong Gyimah, 1986; Armah, 2008)

A shift to increase production would contribute to the national economy through as increase in foreign exchange earnings, an improvement in the GDP of the country as well as an improvement in the balance of payment (Awua, 2002).

As the world's second largest producer and major contributor to the growth of the country, cocoa is highly valued by Ghana. The cocoa industry of Ghana consists of cocoa bean production by smallholder farmers, collection and bagging by Licensed Buying Companies (LBCs), quality assurance by COCOBOD, haulage of cocoa by private hualers, warehousing and other logistic by private companies and COCOBOD, and exports to external buyers by COCOBOD, (Amoah, 2008). Currently, Ghana export about 70% of its raw cocoa beans (COCOBOD, 2009). That is, the farmers grow the cocoa seeds. The pods are then collected, broken and the beans extracted. The beans are then harvested, fermented, dried and bagged for export (International Cocoa Initiative, 2008). This is the process that 70% of cocoa from Africa 19% of cocoa from Asia and Oceania, and 11% of cocoa from the Americas go through (World Cocoa Foundation, 2009).

Other processes on the cocoa value chain include cleaning, roasting and removing of the shell of the bean (International Cocoa Initiative, 2008). The nib in the shell is ground to form a cocoa paste. This paste can be pressed to extract cocoa butter which represents 50% of the cocoa bean. The remaining is the cocoa powder which is typically used for producing cocoa drink, for baking and in the cosmetic industry. It is also used in chocolate , confectionary and other food product (International Cocoa Initiative, 2008).

Together these activities from Cocoa processing which is distinct from production. This processing mostly occurs in industrialized counties in Europe and the United State. Europe, America, Asia and Oceania, and Africa grind approximately 41.1% m 22.7% 19.5% and 16.6% of world cocoa beans respectively (World Cocoa Foundation, 2001).

According to the Global Trade Atlas, as of March 2009, the Netherlands imports cocoa beans to the tune of USD 352,505,018. The United States imports USD 40,785,667, worth of cocoa powder and France imports USD 288,328,480 worth of retail chocolate. These three countries are the highest Importers in the respective categories of cocoa product (World Cocoa Foundation, 2009).

Amongst the regions which grind cocoa, Africa is the only one with a positive growth of 11.7% from 2007/2008 to 2008/2009. Currently of the 16.6% grinding in Africa, Ghana contributes about 25.8% (World Cocoa Foundation, 2010). Based on the various forms of products into which cocoa can be processed, avenue for maximum revenue for Ghana possibly be achieved by Increasing . Thus this thesis aims to investigate the production and benefits of moving up the supply chain for Ghana's cocoa sector.

Specifically the thesis will use a production analysis based on the regional production and revenues from production of cocoa beans to determine whether it would be more beneficial to produce more cocoa bean for export and the rest locally process to help increase the GDP of the nation.

The importance of cocoa to Ghana has been identified by several previous researchers. In Ghana, cocoa has been the backbone of the economy for a century and plays a major role in employment, foreign exchange earnings, government revenue, education, infrastructural development amongst other. (Amoah, 2008).

The International Cocoa Initiative, assert that over 14 million workers produce cocoa, of which 10.5 million are in Africa 95% of the world's cocoa is grown by small scale farmers.

In Ghana, it is estimated that there are about 265,000 cocoa farm owners and roughly 800,000 people involve in cocoa growing and these figures exclude those working in other areas of the industry such as the processing firms, Licensed Buying Companies, chocolate vendors and other (Awua, 2002).

This therefore implies that the demand for cocoa directly affects many African Nations and their citizens livelihoods. Apart from the over reliance of some families on cocoa, a dependency on the revenues from cocoa by the country as a whole could be quite detrimental especially since the price of this commodity is largely determined there was a level of instability inherent in relying on one export commodity for revenues.

Historically, Ghana has shown some over reliance on revenues from cocoa. Aryeetey and Kanbul (2008), noted that Ghana's first president, Kwame Nkrumah, used cocoa revenue as security for loans to establish different state-owned industries. Nkrumah's dependence on cocoa, along with the fall in prices in the late sixties(60s), caused a decline in the growing of the country and resulted in a coup to overthrow him. Sahn (1994), also states that from the introduction of cocoa in the late 19th century till the mid 1970's Ghana dominated the world cocoa market, and to a large extend cocoa dominated Ghana. Ghana generated GH¢678,932,789,754 from cocoa and made the GDP to rise 5%. This clearly shows that if Ghana is able to produce a substantial amount of cocoa there will be a rise in the economy.

1.2 JUSTIFICATION

It is striking to note that despite enjoying the enviable position of producing the highest quality cocoa the world over, the scale of cocoa production in Ghana remain relatively low. Why this and what are the economic underpinnings of this observation? Why does Ghana not produce more cocoa to increase the annual GDP since it is a major contributor. Given the recent ramp up in Ghanaian cocoa production and the decline in Cote D'voire's ability to export cocoa Ghana can produce more cocoa since there is now equal producer price of cocoa.

Ghana exports 80% of the bean production annually and the remaining 20% is processed. But if cocoa production is increased by 40% then our GDP will increase and we will be able to increase the percentage of the locally cocoa beans.

The motivation for the research emerged from a genuine curiosity about the relative profitability of increasing the percentage of Ghana cocoa that is produced in Ghana.

1.3 STATEMENT OF THE PROBLEM

Since independence Ghana has depended mainly on exports of raw cocoa beans for majority of its foreign exchange revenue, although exporting cocoa beans is hardly the only market strategy available to Ghana but since 1980 the production of cocoa in Ghana is still below that of Cote D'ivoire.

The thesis therefore analyses whether the mean difference in cocoa production in the six(6) growing regions in the country is significant and where the difference lies among the regions for stakeholders to check what can be done to close the gap in difference to boost production.

The results of this thesis would provide investors with information that channel their money to other more profitable ventures. Also based on this study, one will be aware of the advantage or disadvantages of encouraging large foreign production companies such as Archer Daniels Midland (ADM) and Cargill into the country. Lastly, this paper will make a contribution to the literature that analyzes the advantages and disadvantages of moving up the vertically integrated agricultural commodity-based production chain.

1.4 RESEARCH OBJECTIVES

The primary aim of this thesis is to identify, analyze and compare the mean differences of the major cocoa production regions in Ghana as to achieve the aim of COCOBOD in the 1.2 million tones target by 2012/13 production year. Specifically, the study's objectives include an assessment of:

1. To identify regions that produce high or low cocoa yield in the country over the forty(40) years.
2. To use multiple comparison test procedures to identify the significant difference in the mean cocoa yield production among the six(6) regions in the country.
3. To account for the variations in the cocoa yield production using random effect model.

1.5 SIGNIFICANCE OF THE STUDY

This study will augment existing knowledge on the continued profitability and sustainability of the cocoa industry in Ghana. Given the ongoing nationwide discussion about whether there is economic justification for production up to 40% of cocoa beans to increase the GDP of Ghana by exporting of raw cocoa beans, this research will give reader insight on the benefit of such a decision.

This study can serve as a springboard for further research on Ghana's cocoa industry. It will be especially useful to farmers who want to expand their scope of cocoa operations or to entrepreneurs who might start cocoa farms. Their awareness of the most profitable stage of cocoa along the supply chain (export of beans, processing and exporting processed cocoa products etc.) would enable them maximize their profit from their operations if they were to expand their farming business. Also the study will be useful for investors and businessman who seek to devote resources or capital in the agricultural sector. Most importantly , the Ghana government and COCOBOD may benefit from this study because awareness of what stage of cocoa production enjoys the most significant margins, will guide plans for expansion of the cocoa industry and will enable them create policies and regulation which would boost the industry.

1.6 SCOPE OF THE STUDY

This study was structured using the six (6) regions where cocoa has been produced in Ghana. The study focused on the analysis the mean, establishing multiple comparison tests and modeling an equation for mean cocoa production in Ghana. The data was obtained from COCOBOD in Accra which covers the period of 1969/70 to 2009/2010 production years.

1.7LIMITATION OF THE STUDY

This research experienced limitations in the area of limited information, limited time period for conducting the research and financial constraints. This research was completed using mainly secondary data which may have some errors in the data collection. Again due to data collection constraint in this country this work was conducted using a small sample size. The above limitations, however, do not render the findings of this research non-reliable and applicable since the researcher carefully managed these limitations to make sure the research objectives were achieved.

1.9 ORGANIZATION OF THE STUDY

The study has been organized into five chapters, the first chapter commences with the introduction, including: Background of the study, statement of the problem, objectives of the study, significance of the study, methodology of the study, scope and limitation of the study. A review of relevant prior literature on the origin of cocoa, its production in the country and the importance of cocoa to the nation constitutes chapter two. The third chapter focuses on the methods used in the study followed by chapter four which constitutes data analysis and summary of the results. The final chapter (five) gives a summary of the research work findings, conclusions and recommendations and it is followed closely with references and appendices.

CHAPTER TWO

LITERATURE REVIEW

This chapter reviews the literature relevant to the theme of this study. First, the history on the origin and the spread of cocoa is reviewed. Secondly, it describes briefly the structure of cocoa production and the role of cocoa in the economy of Ghana is reviewed. Third, the set of literature on the determinants of cocoa output in Ghana is well touched and lastly literature review on statistical method used.

2.1 Origin and Spread of Cocoa cultivation in Ghana.

Cocoa, ***theobroma cacao,*** originated around the headwaters of the Amazon in South America (COCOBOD, 1998). Its cultivation and value spread in ancient times throughout Central and Western Amazonia and northwards to Central America. Large scale cultivation of cocoa was started by the Spanish in the 16th Century in Central America. It spread to British, French and Dutch West indies (specifically, Jamaica, Martinique and Surinam) in the 17th Century and to Brazil in the 18th century. From Brazil, it was taken to Sao Tome and Fernando Po in 1840 (COCOBOD, 2000). From there, it spread to other parts of West Africa, notably the Gold Coast (now Ghana), Nigeria and Cote d'voire. Records show that Dutch missionaries planted cocoa in the coastal areas of Ghana as early as 1815, whiles in 1857, and Basel missionaries also planted cocoa at Aburi (COCOBOD, 2000). However, these did not result in the spread of cocoa cultivation until Tete Quashie, a native of Osu, Accra who travelled to Fernado Po to work as a blacksmith, returned in 1879 with Amelonado cocoa pods and established a farm at Akuapem Mampong. It later spread to other parts of the Eastern, Western, Ashanti, Brong Ahafo and Volta Regions.

The seriousness with which the people of the Gold Coast took cocoa farming was phenomenal. Knap (1920), for instance, observed that the enthusiasm and seriousness with which cocoa was cultivated in the Gold Coast shattered the stereotype image of the "indolent" native, and showed the World that the "native" was capable of building a strong economy by their own initiative and industry. As a result of its high demand in the European and American, it quickly became the main traditional export commodity in Ghana.

2.2 Cocoa Production in Ghana

2.2.1 The Structure of Cocoa Production

The *Amelonado* cocoa was the first cocoa type to be introduced in Ghana. It takes not less than five years to bear fruit. The Amazonian type which takes three to four years to mature was introduced into the country in the 1950's. Almost all the cocoa farms established in the 1960's and 1970s were sown to the Amazonian type. Recently, the hybrid cocoa variety (called "*akokora be di"* in Akan) was introduced. This is high yielding and early maturing. Cocoa is a perennial tree crop with a life-cycle of twenty-five to thirty years. In the initial stage of land cultivation (two to three years from planting depending on the tree variety) it is intercropped with staple food crops like maize, plantain, cassava and cocoyam, which provides shades to the young trees until they grow and form a closed canopy, at which point they are left to and stand alone. Cocoa trees typically take between three to six (depending on the variety)years from planting before they start bearing the first pod, and full production capacity is only reached after ten years from first planting. The ideal climatic zone under which the crop is grown is the tropical rainforest zone.

Cocoa trees grow under shaded conditions with a climate characterized by relatively high temperatures (between 18-32 degrees Celsius) and plentiful rainfall. Cocoa production also depends heavily on the pattern of rainfall; the average distribution of monthly rains throughout the year is more important than the annual total. Annual rainfall in excess of 2500mm may lead to a higher incidence of fungus diseases, the most common known as *phytophtora* pod rot which causes the black pod disease, and the cocoa swollen shoot virus (ICCO, 2000; Wood and Lass, 1985). Cocoa needs deep, well-drained soils adequately supplied with nutrient and moisture and containing little or no coarse materials (Dickson and Benneh, 1985). The cocoa belt in Ghana generally coincides with the semi-deciduous forest zone. Land preparation for the cultivation of cocoa is done in the same way as for foodstuffs. The forest vegetation is first cleared, but with some of the trees left standing. The litter is burnt during the height of the dry season. Traditionally, foodstuffs like cocoyam and plantain are first planted at the start of the rain in March, and the cocoa seedlings are planted among them to shelter under their broad leaves. Cocoa farms only need occasional weeding and brushing to control weeds (personal communication with farmers). Depending on the variety of cocoa seedlings planted, it may take three to six for the tree to bear fruits.

Cocoa trees typically have two harvest seasons in the year, the main crop (which begins in October and ends in March) and the smaller or mid crop season (between May and August.

Harvesting or picking of the ripe cocoa pods starts from about September and may continue till late December or mid-January, depending on the size of the crop (Personal communication with experienced cocoa farmer in Sefwi-Bekwai). It is done by means of a cutlass or a metal hook which is so constructed that it serves the pod neatly from the stem by a thrust or a draw. Labour is mainly supplied by family. The women collect the harvested pods into heaps and carry them in baskets to a spot selected for breaking which is done communally. The men cut open the pots with cutlass and women and children scoop out the wet cocoa with their hands. The beans are fermented for a period of 6-7 days in the wrapped, airtight container made of banana or plantain leaves. Sometimes, it is heaped under the multi-storey canopy of cocoa trees and well covered with broad leaves, usually plantain leaves. The fermented beans are then transferred to raise drying platforms made of sticks and covered with mats of split bamboo. The dried beans are then collected into mini or maxi bags of 30 kgs and 62.5 kgs respectively and are sold to local buying agents who are distributed throughout the cocoa growing regions. They are then weighed, graded and bought at prices fixed by COCOBOD.

2.2.2 The Role of Cocoa in Ghana's Economy

In the role of cocoa in Ghana's future development Breisinger et al (*2007*), describe the recent performance of the sector as an example of what *"favorable external conditions and internal reforms"* can do to renovate traditional exports. Ghana has maintained overtime a leading position among cocoa producing countries, despite the criticism by economic commentators that its continued dependence on traditional export crops might push the economy into the dependency trap from raw commodities (of which cocoa contributes the bulk of the country's foreign exchange earnings together with gold and timber).

Serious concerns also arise over the future sustainability of the sector, as recent research findings clearly indicate that past and present cocoa growth have been driven by land expansion and by the intensive use of labor, rather than by rise in land productivity (Gockowaski, 2007 and Vigneri,2005).

Cocoa contributes about 70 per cent of annual income of small scale farmers and stakeholders like licensed cocoa buyers (LCBs) also depend largely on their products for market, employment and income (Asamoah and Baah, 2002). Knudson (2007) shows that income from cocoa is still the determining factor for most households' income and thereby for the demand for non-farm foods and investment in the non-farm sector. Many researchers have been able to show that small farms are desirable not only because they provide a source of reducing rural unemployment, but also because they provide a more equitable distribution of income as well as an effective demand structure for other sectors of the economy (Bravo-Ureta and Evenson,1994 and Dorner, 1975). The cocoa growing industry also provides employment to many Ghanaians. It occupies well over one-third of Ghana's cultivated land and well over 55 per cent of farm families are directly and indirectly engaged in the production of cocoa (Dickson and Benneh, 1995). The cocoa sector in Ghana employs over 800,000 smallholder farm families. The number of cocoa farm owners is estimated at 350,000. Moreover, in 1960, when population census was taken, there were as many as 552,350 people directly engaged in the industry, including 312,510 cocoa farmers, 500,080 caretakers, 908,040 family workers and 68,920 hired laborers (Dickson and Benneh, 1995). Moreover, there were numerous cocoa purchasing clerks, drivers and others involved in the purchase and shipping of cocoa to the European and American markets. In addition other stakeholders like chemical companies, input distributors and licensed cocoa buying companies also depend largely on cocoa for markets for their products employment and income.

The fact that agriculture (including cocoa) is the driving force of the economy simply means a decline in this sector is likely to lead to a decline in the growth of the economy as a whole. Ghana produced one-seventh of World cocoa in the late 1980s and early 1990s as compared to one-third in 1965. This tremendous decline was partly attributed to inadequate credit supply, inappropriate control of disease and pest and poor macroeconomic policy, just to mention few (Opoku, 2003). Between 1986 and 1992, a decline in cocoa production led to a decline in foreign exchange earnings by 41 per cent; that's, decreasing from US $503.3 million to US $302.5million (Compton Interactive Information Guide, 1995). This led to a high nominal producer price to boost production, but that was not enough to sustain increased real producer price of cocoa and cocoa exports due to huge inflation (Compton Interactive Information Guide, 1995). Between 1970 and 1982, Ghana Gross Domestic Product (GDP) declined by 0.5 per cent per annum, real export earnings fell from 21 per cent of GDP of 1970 to 4.0 per cent in 1989 (UNCTAD, 1990). This, among others factors, led to the launching of the Economic Recovery Programme (ERP) in April 1983 as one of the government interventions to reshape the economy. This programme, inter alia, was to increase the incentives for food production, raw material production and traditional export crop production which cocoa was an important component.

Sales of cocoa beans have been one of the major foreign exchange earners to Ghana throughout the years. In 2002, cocoa made up for 22.4 per cent (463 million US $) of the total foreign exchange earnings (ISSER, 2003). Cocoa constituted 63% of the foreign export earnings from the agriculture sector (ISSER, 2003). Cocoa is the only traditional export commodity whose export is taxed; in 1998, it contributed 14.5 per cent of total tax revenue in the country (ISSER, 2000). The total export receipts from cocoa (beans and products) in 2002 amounted to US$463.4 Million compared to US$381.1 million in 2001, representing an increase of 17.8 per cent (ISSER, 2002). The cocoa sub-sector exhibited the most impressive performance in recent time. For instance, the cocoa sector grew at an outstanding rate of 16.4 per cent in 2002. This has been attributed to both increase in cocoa output and relatively better border price for the commodity (ISSER, 2003).

2.4 **Determinants of Cocoa Production**

Although models of cocoa supply in Ghana are found more frequently in the literature than models of other perennials in the economics literature, the sum total of models of perennials in general including cocoa models remains unimpressive (Bulir, 2002). The biological lag between the planting decision date and output date presents unique challenges for econometric modeling not only for cocoa, but also for all perennials. Empirical problems also arise because of incomplete, unrecorded or missing data pertaining to plantings, removals and re-planting, yield variations and yield composition (King et al, 1985). The cocoa supply modeling literature has therefore evolved as different analysts have tried to obtain more accurate forecast models by taking into account not only the lag but also other exogenous factors that affect output; for example, cocoa output price instability, cocoa production variability, probably caused by bad weather and also the availability of inputs into production (or rather the lack thereof) have all received considerable attention in the literature (King et al, 1985 as cited in Armah, 2008).

According to Bluir (2002), studies on cocoa modeling can be divided into three broad categories. First, some studies model the supply of cocoa as a "technological" function of the stock of cocoa trees and fertilization effects resulting in long-run or a short-run function that takes into account price and weather shocks,. Second, a traditional partial-Adjustment supply model has been used with properly defined elasticity of domestic producer prices. Finally, few studies have estimated the supply response to changes in producer prices in neighboring countries. These studies have generally found that smuggling explains supply fluctuations better than most other variables.

The second and the largest group of empirical studies have concentrated on the traditional partial-adjustment model using several domestically determined explanatory variables (Abbey and Clark, 1974; Yeung et al., 1979; Berthelemy and Morrison, 1987 and Stryker et al, 1990). In these studies, the estimated equations and the results of those estimates are similar. As a representative example, Stryker *et al.* (1990) have regressed the actual production on its lagged value, an estimate of cocoa production capacity, producer prices of cocoa, and the producer prices of competing food crops. The estimated own short-run and long-run producer price elasticities were 0.22 and 0.62, respectively, and the cross-price elasticities estimated at -0.14 and -0.40 respectively.

The third group of authors focused on the price incentives to smuggle to explain why the officially recorded cocoa production stayed for several years above or below its estimated production capacity. These authors realize that cocoa is a ***Golden Cash Crop*** that can be easily smuggled, because the boarders contribution to the first group (technological capacity model). As a first step, he estimated a long-run production capacity for Ghana based on tree yields among several variables measuring the chemical spraying of cocoa trees that had a built-in ratchet effect (Bateman, 1974). As a second step, his short-run function included the previously estimated production capacity, real producer price and rainfall variables. Both equations were estimated separately for the three major cocoa producing regions in Ghana, and the short-term price elasticities of supply were found to be of similar magnitudes, ranging between 0.14 and 0.22 with Cote d'Ivoire and Togo are practically unguarded. As early as 1982, Akiyama and Ducan (1982) regressed cocoa output on real prices (both in first-order differences) and a rainfall variable; in addition, their equation included three variable lagged one year: cocoa output, real producer prices, and the Ghana-Cote d'Ivoire price differential (all in level). Both short-run and long-run domestic producer price elasticities were low and statistically insignificant. However, their models showed the strong impact of price development in Cote d'Ivoire: raising the price differential by 1 percent lowered the Ghanaian supply of cocoa by one-quarter of 1 percent. In order words, the official sales of cocoa to COCOBOD/Ghana might have fluctuated because of smuggling rather than changes in cocoa output growth. Fosu (1992) supported these findings; he estimated the short-term elasticity of Ghana's cocoa export with respect to the Ghana-Cote d'Ivoire price differential at about 0.17. May (1985), in estimating the regional motivation to smuggle cocoa to neighboring countries, found that as much as 50 percent of the crop in some regions may have been smuggled either to Cote d'Ivoire or to Togo. As a result, he found that virtually all new cocoa plantings in Ghana in the 1970s and 1980s were made in areas adjacent to Cote d'Ivoire and Togo in order to minimize the cost of transporting smuggled cocoa. Azam and Besley (1989) formulated and tested a general equilibrium model of Ghana's economy that features parallel foreign exchange and consumer good markets, and cocoa smuggling.

2.4.1 **Control of Diseases and Pests of cocoa**

The high incidence of pest and disease infestation is considered by many farmers to be the major cause for low cocoa yields (Nyanteng, 1980). Three major diseases and pest of economic significance exist: (I) swollen shoot caused by virus, (ii) black pod caused by fungus and (iii) capsid, which feed on plant tissues (shoot and pods), eventually killing them. Many diseases affect cocoa on the field, some of them are *phytophtora* black pod disease, *phytophtora* canker, *phytophtora* seedling blight, *Theilaviopsis* pod rot, cocoa swollen shoot virus (CSSSV) disease, *Cherelle wilt, charcoal* pod rot and *Collar crack* disease (Adegbola, 1972).

Black pod disease probably appeared as soon as cocoa was introduced in Ghana and it is considered to be the most destructive among all cocoa diseases which attack the developing cocoa pod. It is caused by soil-borne fungus *phytophtora* and is most prevalent during the wet season. The disease is worse in the areas of heavy rainfall. The disease can cause severe damage, rotting both small and large pods. Coupons, seedlings (in the nursery) and leaves of cocoa can be attacked and destroyed under conditions of long periods of cool and rainy weather. Losses of cocoa yields due to black pod disease vary from place to place and from variety to variety. Adegbola (1972) put the average to 40 percent over several parts of West Africa and up to 90 percent in certain places in Nigeria. Crop loss due to this disease was estimated at about 29 per cent in the 1950s and 1960s (Wharton, 1962). Deduction from analysis of data from the Cocoa Research Institute of Nigeria (CRIN) indicates that pod loss due to black pod diseases infection varies with variety of cocoa. The average percent pod loss over the years 1962-1993 was 7.56 for Amazon 1, 6.56 for Amazon II, 7.01 for Amazon III and 13.03 for Amelonado. This is not quite different from the rest of West African countries (Tijani, 2005). In Ghana, the black pod disease is caused by *two phythophthora* species: *P. palmivora* and *megkarya* (Opoku et al, 1999). Generally, losses due to *P. megakarya* range from 60-80 per cent in newly affected farms to about 100 per cent in old affected farms in the black pod season (May to mid June). Losses for P. palmivora are estimated at 4.9 per cent to 19 per cent (Dakwa 1984). This deadly disease, through yield reduction, also reduces farmer's revenue and the country's export earnings. The recommended method of control was to remove the affected pods and also to harvest the matured pods at short intervals.

However, harvesting at short intervals does not meet the requirements for proper fermentation to obtain quality dry beans. Farmers therefore, prefer harvesting at long intervals, which unfortunately promotes a high incidence of the disease. Since the mid-1980s fungicides have been recommended for the control of the disease (Nyanteng, 1980). Babcock et al. (1992) noted that those yield loss could be reduced through the use of chemical control agents (synthetic pesticides) which have been favored because of their effectiveness (it diminishes with time in many cases), their relative shelf life (when properly stored), and the ease with which they can be transported, stored and applied. It should however not be forgotten that cocoa farm families spent huge amount of money in the procurement and application (labour cost) of these chemicals thus draining the income of these poor small-scale farmers.

Oluyole et al. (2008) estimates the determinants of the occurrence of black pod disease of cocoa. He uses the probit analysis approach to determine the influence of some explanatory variable such as availability of fungicides, price of fungicides, price of cocoa beans, and labour availability among other things. The parameters of the probit model were estimated by maximum likelihood estimation rather than by Ordinary Least Square. Price of fungicides was a significant determinant of the probability of cocoa farm having black disease ($P < 0.05$). This simply means that the higher the price of fungicides, the higher the probability of occurrence of black pod disease. In addition, price of cocoa beans was significant determinants ($P < 0.1$) of the probability of the occurrence of the disease. That is, the higher the price of the cocoa beans, the lower the probability of the occurrence of the disease. It can therefore be inferred from Oluyole's analysis that increase in the producer price of cocoa can help reduce the probability of occurrence of black pod disease. However, none of these studies focus on the effects of fungicide application on output and how efficient the chemical is used in controlling the disease, hence this study bridge that gap of knowledge.

Capsids which causes the swollen shoot disease were first identified as serious cocoa pests in the early part of the cocoa beans industry's history, 1910 (Asomaning, 1971). In the mid-1950s, it was estimated that about 50 per cent of the total cocoa area was severely damaged by capsids, serious attempts to control the insects were made in the late 1950s and directed by government who organized two mass spraying campaigns. The first covered only the western part of the Ashanti Region (Nyanteng, 1980). Following mass spraying campaigns, responsibility of capsid control was then transferred to farmers. It was reported that by early 1960s, capsid damage had been brought under control (Addo et al, 1979). Since then, the supply of insecticides and spraying machines has not been adequate to met estimated requirements for effective spraying of all cocoa farms. A country-wide mass spraying campaign was designed and implemented to cover only the Ashanti and Brong Ahafo

Regions during the 1978/79 season; it was subsequently terminated without achieving the target. In the 1970s, capsid damaged accounted for an estimated 50,000 to 75,000 tones in the production loss each year (World Bank, 1980).

In respect of the application of fungicides against black pod disease and insecticides against swollen shoot disease, various suggestions have been made. Opeke (1987) suggested early spraying in the season and application repeated every three weeks until rains ceased. Cocoa Research Institute of Ghana also recommends an average of seven to eight times of spraying fungicides per season and three to four times of insecticides spraying per cocoa season.

The foregoing presupposes that chemical control of cocoa diseases (mainly black pod and swollen shoot diseases) is feasible, acceptable and politically advantageous (Norton, 1993). However, in the face of escalating costs of agricultural input (insecticides and fungicides), economic desirability appears very questionable. Nyanteng (1980), found the following to be some of the reasons for farmer's inability to spray their farms as often as recommended: lack of adequate quantities of insecticides, lack of funds to buy insecticides and unavailability of motorized spraying machines. It follows that, given that these constraints persist, an increase in the usage of insecticides resulting from low cost (subsidization) of insects would increase output per hectare and hence increase farmers revenue. The studies on cocoa insects and diseases and their control reveal that there is a knowledge gap concerning the magnitudes and the directions of the effects of the application of fungicides and insecticides as well as their frequencies of application on the output of cocoa in Ghana's cocoa bean industry.

2.4.2 Fertilizer Application

Low soil fertility has been identified as one of the major causes of decline in yield of cocoa. The significance of fertilizer in ameliorating this problem will go a long way to boost cocoa production. Replacement of soil nutrients that are being mined through cocoa pod harvest annually cannot do without application of fertilizer. Adequate application of fertilizer has been found to increase agricultural output. Traditionally, Ghana's cocoa was grown with minimum purchased inputs, although it has long been recognized that soil nutrients reserves would become exhausted (Charter, 1953). Recently, Appiah, et al (1997) argues that soil nutrients availability has indeed become limiting to cocoa yields. Appiah, et al. (1997) reported a doubling of yields in Ghana from the applications of 4.94bags of triple superphosphate and 2.47bags of muriate of potash per hectare over 4 years. According to Olson (1970), fertilizer could increase food production by at least 50 per cent. Opeyemi et al. (2005) in their recent work noted that, an effective use of fertilizer on cocoa would help not only to improve yield but also has the advantages of profitability, product quality and environmental protection. FAO (1987) noted that tremendous increase in fertilizer use has the highest potential of increasing productivity.

Ogunlade et al. (2009), use regression analysis to assess the determinants of the quantity of fertilizer usage of cocoa production. The quantity of fertilizer used was regressed on explanatory variables like farm size, fertilizer availability, and rate of fertilizer application and the price of fertilizer. They showed that the farm size as well as the price of fertilizer was much more critical in determining the quantity of fertilizer to be used. However, the fertilizer availability as well as rate of fertilizer application has no influence on the quantity of fertilizer used by cocoa farmers. However these authors did not quantify the effects of fertilizer quantity and its usage on annual cocoa production and, hence this work seeks to fill that gap.

2.4.3 Rainfall

Cocoa just-like any other crop, is responsive to rainfall and highly susceptible to drought and the pattern of cropping of cocoa is correlated to rainfall distribution. There is a significant correlation between cocoa output yield and amount of rainfall over varying interval prior to harvesting. In Ghana, a year with high rainfall is followed by a year with larger crop output, though the correlations not applicable in all years (Brew, 1991). Ali (1969) reported both positive and negative correlations between rainfalls in certain months with the mean of yield crop in Ghana. The annual total rainfall in the cocoa growing regions of Ghana is less than 2000mm. The rainfall distribution is bi-modal from April to July and September to November. Cocoa as a tropical crop can only be profitably grown under temperature between 30-32C mean max and 18-21C mean minimum and absolute minimum of 10C, Temperature has been related to light use efficiency with temperatures below 24˚ C having a decreasing effect on the light saturated photosynthesis, (Wood and Lass, 1985).

Anim-Kwapong and Frimpong (2008) estimated the impact of climate changes on the supply of dry cocoa beans. Their work sought to determine the effect of changes in total annual rainfall, total rainfall in the two driest months and sunshine duration.

They used multiple regression analysis to show that over 60% of variation in dry cocoa beans could be explained by the combination of the preceding total annual rainfall, total rainfall in the two driest months and the total sunshine duration. Oyekale et al. (2009) also showed that about 82 percent of cocoa farmers in Nigeria depend heavily on rainfall and could be more in the rest of West African countries. They estimated the impact of climate change on the production of cocoa. It was stated that, the main climate was rainfall and has a very significant impact on cocoa growth. Rainfall failure therefore has the ability to increase the cost of controlling diseases and pest and reduce the quality of the cocoa beans. Excess cost and reduce quality were significant at 1 percent and 5 per cent respectively.

2.4.4 The Producer price of Cocoa

One of the key economic policies in Ghana each year is the setting of the producer price of cocoa. Farmers, as any other rationale producers, respond to price by changing the intensity with which they tend their farms. If prices are not enough to cover their normal average variable cost including maintenance, the farmer's first response will be to reduce maintenance of the farm and stop new planting activities. If prices do not even cover harvesting, fermenting and drying, then harvesting is most likely to cease. Conversely, if prices cover or exceed variable cost, farmers will intensify farm. The short-term price elasticity of supply is estimated at 0.3 and the prices elasticity of production for period 5 to 10 years later are 0.9 and 1.8 respectively (COCOBOD, 1998).This means that a 10% increase in real prices will result in 3% increase in production in the short term. In the longer term increases in production resulting from new plantings will be about 18% higher after 10 years (COCOBOD, 1998)

The volume of Ghana's cocoa exports has expanded significantly in the last several years after many years of decline followed by a mediocre performance recovery (ICCO 2005, IMF Country Report, 1995). Not surprisingly, cocoa prices paid to Ghanaian cocoa farmers have also appreciated both in nominal and real terms; The nominal price per bag of cocoa beans paid to farmers by Ghana Cocoa Marketing Board (COCOBOD) which was GH¢7 in 1995, topped GH¢90 by 2004, representing an astronomical increase of 1186% although after exchange rate effects and inflation are accounted for this increase is less impressive (Ministry of Agriculture, 2005). To explain the severe contraction in Ghanaian cocoa supply from 1960's to the 1995s (a 60% decline) Bulir (2003) appealed to the reversal in price-incentive to smuggle Ghana cocoa to Cote d'voire using co-integration model and a single equation error correction model. He explained that distortionary effect of domestic taxes in Ghana widened the gap between the Cote d'voire and Ghanaian domestic prices, and ultimately created incentives to smuggle Ghana cocoa to the CIV (Bulir, 2003).

Bulir argued that the monopoly position of Cote d voire enabled that country to pay better domestic prices to its farmers. Rational Ghanaian farmers therefore smuggled their cocoa to Cote d'voire when the expected gain from smuggling Ghana cocoa to Cote d'voire outweighed the transportation and transaction costs that this risky adventure entails. Armah (2008) also showed that the smuggling incentive was statistically significant at 5% and that the international cocoa price is positively statistically significantly related to cocoa supply in the long run while the cocoa producer price correlate to supply response in the short run.

So as the producer price of cocoa increases, Ghanaian cocoa farmers responded by supplying more cocoa both in the short and long run.

Some studies have estimated the effects of input supply of cocoa and its use (Bateman, 1994; Nyanteng, 1990 and Okyere, 1989). These authors looked at how increases in the real producer price of cocoa encourage the establishment of new cocoa farms, rehabilitation, replanting and maintenance of existing cocoa farms. They did not however quantify the effects of the trends of input cost (for example, fertilizer cost and insecticides cost) on the supply of cocoa beans with regards to resource use efficiency. Fosu (1992) indicated that most of the factors postulated to influence cocoa export supply in Ghana are directly or indirectly related to the real exchange of the domestic currency. He further stressed that it is in fact a major factor in the decline of cocoa exports.

Bateman (1973) estimated the effects of the domestic real cocoa producer price and weather on cocoa supply. His work sought to determine the effects of changes in producer price, insecticides usage and government extension programmes on cocoa yield in Ghana. He first specified cocoa base capacity as a function of past planting, tree yields and insecticides (gammalin 20) application. The average capacity estimate from this function was then introduced into a short run supply fluctuations equation. The equation allowed deviation of prices and rainfall to either increase or decrease relative to average capacity.

2.4.5 Labour and Other Socio-economic Factors

According to pilot survey conducted by Ministry of Employment and Manpower Development in 2006, the age of range of adult workers in cocoa farms in the cocoa growing regions in Ghana is between 18 years and 70 years, but most of the workers (76.3 per cent) belonged to the younger age grouping of 18 years to 35 years, indicating that most of the workers were relatively young. Labour constraints constitute an important determinant of cocoa supply (see for instance, Okali and Rouke, 1974; Manu, 1974 and Robertson, 1987). Studies on labour constraints relating to cocoa production have tended to concentrate on ways to improve the standard of living of cocoa producers, since family labour was largely used in the cocoa production process in Ghana. Other labour issues addressed included how labour shortages affected cocoa output and the importance of labour in cocoa production as well as the advantage effects of the deportation of illegal immigrants who were a source of their labour, the organization of labourers into society, social security and insurance scheme, making land available and improving health facilities in cocoa communities. Surveys conducted by Ministry of Finance (1998) on cocoa farms show that about 25% of current cocoa tree stocks are over 30 years old. Behrman (1968) showed that if cocoa were allowed to reach maturity, there would be large output response to the cocoa real producer price. He indicated that the estimated average long-run elasticity was about 0.9. The long run elasticity is the response after newly planted trees have come into full bearing and all other adjustments have been made.

2.5 The Concept of Efficiency in Agricultural Production

The concept of efficiency is defined as the index of the ratio of the value of total farm output to the value of the total inputs used in farm production (Olayide and Heady, 1982).

The main aim of resource use efficiency is to find ways of increasing output per unit of input and attaining desirable inter-firm, intra-firm and inter-sector transfer of production resources in order to provide the means of raising our economic level of living. Three main types of efficiency are identified in literature: These are technical efficiency, allocative efficiency and economic efficiency (Olayide and Heady, 1982; Farrel, 1957). Technical efficiency is the ability of a firm to use a minimum quantity of inputs under a given technology to produce a given level of output. Allocative efficiency is defined at the firm's ability in achieving the best combination of different inputs in producing a specific level of output considering the relative prices of these inputs. Economic efficiency is a product of technical and allocative efficiency (Olayide and HEADY, 1982). In other words, the efficiency of a firm is its success in producing as large an amount of output as possible from given sets of inputs. Maximum efficiency of a firm is attained when it becomes impossible to reshuffle a given resource combination without decreasing the total output. Since the seminal work of Farrell in 1957, several empirical studies have been conducted on farm efficiency.However, a variety of statistical tools for determining or analyzing resource use efficiency have been identified by many economists and researchers. Hawksworth (1984) indicated that human resources could be studied through the use of descriptive statistics, questionnaire, surveys and in-depth researches. Kay (1987) also stated that measurement of land efficiency is in terms of yield per hectare of land while capital efficiency, for instance, tractor efficiency, is determined in terms of input and labour power for productive man work unit and could be used to measure the relationship between capitals.

Adesinmi (1981) also identified three major methods of measuring labour efficiency. These include labour efficiency determination in terms of output and amount of labour used, value of the total output and total wage bill as well as total wage bill cum cropped hectarages.

However, uption and Anthonio (1975) also affirmed that labour efficiency could be improved by spreading the labour needs more evenly throughout the year. Therefore, efficiency of labour resource utilization involves optimal utilization of time efficiency profile and work load. It is worth noting that labour supply at busy period could be increased by using communal labour, share cropping or by hired labour.

Measures of efficiency have been classified into three categories namely: deterministic parametric estimation, non-parametric mathematical programming and the stochastic parametric estimation. There are two non-parametric measures of efficiency. The first, based on the work of Chava and Aliber (1983) and Chava and Cox (1988) evaluates efficiency based on the neoclassical theories of consistency, restriction of production form, recoverability and extrapolation without maintaining any hypothesis of functional form. The second, first used by Farrell (1955) decomposed efficiency into technical and allocative. Fare et al. (1985) extended Farrell's method by relating the restrictive assumption of constant returns to scale and of strong disposability of inputs (Llewelyn and Williams, 1996; Udoh and Akntola, 2001). Several approaches, which fall under the two broad groups of parametric and non-parametric methods, have been used in empirical studies of farm efficiency. These include the production functions, programming techniques and recently, the efficiency frontier. Several empirical applications have followed the stochastic frontier specification. These studies are basically based o Cobb-Douglas function and transcendental logarithmic (translog) functions that could be specified either as production or cost function (Udoh and Akintola, 2001).

The first application of the stochastic model to farm level data was by Battese and Corra (1977) who estimated deterministic and stochastic Cobb-Douglas production frontiers for the grazing industry in Australia. Studies relating to allocative efficiency in most parts of African agriculture can be classified into two categories depending on weather a direct (primal) or indirect (dual) method is used. In the primal approach, the production function, in most cases Cobb-Douglas, is directly estimated by OLS technique. After obtaining the parameter estimates, marginal product (MP) of each endogenous input is calculated. The presence of allocative efficiency is then tested by equating the value of MP of inputs with their respective prices (Akinwumi, 1970; Ogunfowora et al., 1975; Umoh and Yusuf, 1999). The dual approach involves estimating the profit function along with the input share (in profit) equation derived from Hoteling's lemma (Udoh, 1999 and Umoh, 2003).

Several studies have sought to estimate Technical Efficiency (TE) and its potential determinants in the Agriculture sector, Tchale (2009; Chirwa (2007); Heshmati and Mulugeta (1996); Helfand (2004); Chomitz and Thomas (2001); Shanmugam and Venkataramani (2006). Binam et al. (2004) in examining the factors influencing technical efficiency of groundnut and maize famers in Cameroon observed an average efficiency of 73% and 77% for the two crops, after controlling for environmental effects. In the cocoa sub-sector, Amos (2007) estimated the productivity and technical efficiency of small holder farmers in Nigeria cocoa industry by employing the famous Cobb-Douglas production frontier. The results of his analysis showed that the efficiency of Nigerian cocoa farmers ranged between 11 per cent and 91 percent with mean efficiency of 72 percent and that there is a scope of increasing cocoa production by about 28% in the short run.

He indicated that the age of farmers, their level of education and family size are major factors contributing significantly to the farmers' efficiency level; while ages of farmers reduce the efficiency level, education and family size increase their efficiency level. Binam et al; (2008) employed stochastic frontier metaproduction to estimate the technical gap and efficiency gap in cocoa production in West and Central Africa. The cocoa producing countries studied were Ghana, Cote'dvoire, Nigeria and Cameroun. For the studied countries, he estimated that, the technical efficiency scores ranged from 0.44 to 0.74, with a weighted average of about 0.61, including that the cocoa sector in West and Central Africa produces on average, only 61 percent of the potential output given the technology available in each country. However, he showed that, Nigeria is the relatively most efficient country with a mean technical efficiency of 0.74 while Ghana is the least efficient country with an average efficiency of 0.44. Binam et al. indicated that imperfect competition, financial constraints etc., may cause a farmer not to be operating at optimal level. However, these studies fail to study how efficient cocoa farmers are in allocating the resources available to them.

2.6 Analysis Of Variance

In statistics, analysis of variance (ANOVA) is a collection of statistical models, and their associated procedures, in which the observed variance in a particular variable is partitioned into components attributable to different sources of variation. In its simplest form, ANOVA provides a statistical test of whether or not the means of several factors (qualitative variables) are all equal, and therefore generalizes t-test to more than two groups. ANOVA is helpful because it possesses an advantage over a two sample t-test. Doing multiple two-sample t-tests would result in an increased chance of committing a type I error. For this reason, ANOVA becomes a useful tool in comparing two, three or more factor level means. Analysis of variances (ANOVA) allows researchers to test for differences in the means of several different groups or populations. ANOVA tests the null hypothesis that the means for all the factor levels are equal. In order to test this hypothesis, an F statistics is calculated which compares the variation among the groups within the groups.

2.6.1 Assumptions Of Anova

There are several approaches to the analysis of variance. However, all approaches use a linear model that relates the response to the treatments and blocks as in Design of experiment. Even when the statistical model is nonlinear, it can be approximated by a linear model for which an analysis of variance may be appropriate. The following are assumptions about the probability distribution of the responses;

- Independence of cases- this is an assumption of the model that simplifies the statistical analysis.
- Normality – the distributions of the random errors are normal. The errors are independently, identically and normally distributed for fixed effects models, that is, that the errors are independent and
- Equality (or "homogeneity") of variances, called homoscedasticity – the variance of factor levels should be the same. Model–based approaches usually assume that the variance is constant. The constant variance property also appears in the randomization (design-based) analysis of randomized experiments, where it is a necessary consequence of the randomized design and the assumption of unit treatment additivity

(Hinkelmann and Kempthorne 2008): if the responses of a randomized balanced experiment fail to have constant variance, the assumption of unit treatment additivity is necessarily violated.

If the assumptions of the model are not met, the results of the analysis may be inaccurate. The visualization for ANOVA is designed to provide information on the validity of these assumptions, as well as information about the significance of effects. It includes a residual plot which is useful for checking if the errors are normally distributed. Also, violation of the homogeneity of variance may be detected by examining the visualizations box plot or residual plot. The partial regression plot provides information about the significance of effects, and the profile plot shows information about the specific levels of an effects.

2.6.2 Types Of Anova Models:

There are three classes of ANOVA models:

i. Fixed-effects models assume that the data were generated from normal populations which may differ only in their means.

ii. Random effects models assume that the data describe a hierarchy of different populations whose differences are constrained by the hierarchy.

iii. Mixed-effect models describe the situations where both fixed and random effects are present.

2.6.3 Fixed-effects models

A fixed-effects model is a statistical model that represents the observed quantities in terms of explanatory variables that are treated as if the quantities were non-random, that is the fixed-effects model of analysis of variance applies to situations in which the experimenter applies one or more treatments to the subjects of the experiment to see if the response variable values change. this allows the experimenter to estimate the ranges of response variable values that the treatment would generate in the population as a whole. This is in contrast to random effects models and mixed models in which either all or some of the explanatory variables are treated as if they arise from the random causes. Often the same structure of model, which is usually a linear regression model, can be treated as any of the three types depending on the analyst's viewpoint, although there may be a natural choice in any given situation.

In panel data analysis, the term fixed effects estimator also known as the within estimator is used to refer to an estimator for the coefficients in the regression model. If we assume fixed effects, we impose time independent effects for each entity that are possibly correlated with the regressors.

2.6.4 Random-effects models

Random effect(s) models are used when the treatments are not fixed. This occurs when the various factor levels are sampled from a larger population. Because the levels themselves are random variables, some assumptions and the method of contrasting the treatments differ from ANOVA model Random effect(s) model, also called a variance components model is a kind of hierarchical linear model. It assumes that the dataset being analyzed consists of a hierarchy of different populations whose differences relate to that hierarchy. In econometrics, random effects models are used in the analysis of hierarchical or panel data when one assumes no fixed effects (i.e. no individual effects). The fixed effects model is a special case of the random effects model. Note that this is not the case in biostaticians call both the"fixed" and "random" effects.

2.6.5 Mixed Effect Model

Mixed models were developed to handle clustered data and have been a topic of increasing interest in Statistics for the past forty years. Clustered data can be loosely defined as data in which the observations are grouped into disjoint classes, called clusters, according to some classification criterion. Example of clustered data include split-plot designs in which the observations pertaining to the same block form a cluster and repeated measures data in which several observations are made sequentially on the same individual (cluster).

Observations in the same cluster usually cannot be considered independent and mixed effect models constitute a convenient tool for modeling cluster dependence. In these models the response is assumed error term. Observations within the same cluster share common random effects and are therefore statistically dependent.

The parameters in a mixed effects model can be classified into two types: fixed effects, associated with the average effect of predictors on the response, and variance-covariance components, associated with the covariance structure of the random effects and of the error term. In many practical applications estimates of the random effects are also of interest.

Several estimation methods have been proposed for mixed effects models and through maximum likelihood and restricted maximum likelihood (Harville, 1974) are generally adopted for linear mixed effects models (Longford, 1993) , there is an ongoing debate in the statistical literature about estimation methods for nonlinear mixed effects models.

2.6.6 **Linear Mixed Effects Models**

Linear mixed effects models in which both the fixed and the random effects contribute linearly to the response function. The general form of such models is

$$y = X\beta + Zb + \epsilon \qquad (2.1)$$

Where y is the response vector, $\boldsymbol{X}$and $\boldsymbol{Z}$ are the design matrices corresponding to the fixed and random effects respectively, $\boldsymbol{\beta}$ is the fixed effect vector**,** $\boldsymbol{b}$ is the random effects vector, and $\boldsymbol{\epsilon}$ is the error vector. It is assumed that $\boldsymbol{b} \sim N(0, D)$ and $\epsilon \sim N(0, \Lambda)$, with $\boldsymbol{b}$ independent of $\boldsymbol{\epsilon}$**.**

Variance components models (Searle, Casella and McCulloch, 1992), mixed effects ANOVA models (Miller, 1977), and linear models for longitudinal data (Laird and Ware 1982) are all special cases of model (1.2.1). the linear mixed effect model (2.1) .

Maximum likelihood (ML) and restricted maximum likelihood (RML) are the most common estimation methods used for linear mixed effects models. The derivation of (R)ML estimates constitutes a rather complex nonlinear optimization problem that only became feasible when fast computers became available. This optimization is usually done using the EM algorithm (Dempster, Laird and Rubin, 1977) or Newton-Raphson methods (Thisted, 1988), but the Latter seems to be more efficient than the former (Lindstrom and Bates, 1988). No closed form expressions are available for the distribution of (R)ML estimates and inference usually has to rely on asymptotic results. The classical asymptotic theory available for MLestimates (Lehmann, 1983) cannot be applied to linear mixed effect models, since the observations are not independent. Miller (1977) derived the asymptotic distributuin of ML estimates for mixed effects ANOVA models, following the work by Hartley and Rao (1967), but these results has not been extended to the more general linear mixed effect model (2.1) under quite general regularity conditions. We also derive the asymptotic distribution of ML and RML estimates of the variance-covariance components in (2.1) for a large class of reparameterization of the variance-covariance matrix of the random effects, that encompasses most cases of practical interest.

2.6.7 **Nonlinear Mixed Effects Models**

Nonlinear Mixed Effects Models are mixed effects models in which some of the fixed and/ or random effects occur nonlinearly in the response function. Several different formulations of nonlinear mixed effects models are available in the literature; we will adopt here the model proposed by Linndstrom and Bates (1990), given by

$$y = f(\phi, X) + \epsilon, \qquad (2.3)$$

Where $\phi = A\beta + Bb$

Where y is the response vector, f is a general nonlinear function, $\boldsymbol{\phi}$ is a mixed effects parameter vector that is expressed as a linear function of the fixed effects and the random effects $\boldsymbol{b}$, $\boldsymbol{X}$ is a matrix of covariates, $\boldsymbol{\epsilon}$ is the error vector, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are the design matrices for the fixed and random effects respectively. As in the linear mixed effects model (2.1) it is assumed that $\boldsymbol{b} \sim N(0, D)$ and $\epsilon \sim N(0, \Lambda)$, with $\boldsymbol{b}$ independent of $\boldsymbol{\epsilon}$.

By far the most common application of model is for repeated measures data. Different estimation methods have been proposed for the parameters in the non linear mixed effects models and there is an ongoing debate in the literature about the most adequate method(s) (Davidan and Giltinan, 1999). One of the reasons for this variety of estimation methods is related to the numerical complexity involved in the derivation of (R)ML estimates in the nonlinear mixed effects model. This complexity is due to the fact that the likelihood functions in the non linear mixed effect model, which is based on the marginal distribution of $\boldsymbol{y}$, does not usually have a close form expression. Different approximations to the log likelihood in (2.3) have been proposed to try to circumvent this problem (Lindstrom and Bates, 1990;n Vonesh and Carter, 1992; Davidian and Gallant, 1993).

As in the linear mixed effects model, the distribution of the (R)ML estimates cannot be determined explicitly. Asymptotic results for these estimates have not been established.

2.6.8 Multiple Comparison Test (Post Hoc Tests)

The analysis of variance (ANOVA) is a powerful procedure for testing the homogeneity of a set of means. However when the null hypothesis rejected in ANOVA and accept the stated alternative that the means are not equal, the ANOVA will not be able to identify which means are different.

With only two groups of observations we could compare the two groups using a t-test. When we have more than two groups, it is inappropriate to simply compare each pair using a t-test because of the problem of multiple testing. The appropriate procedure to do the analysis is to use a one-way analysis of variance (ANOVA) to evaluate whether there is any evidence that the means of the populations differ, we might than be interested in investigating which of the means are different. In this case we employ multiple comparison test, but when multiple comparison is done there is the likelihood of committing at least one type I error. In this case ANOVA and multiple comparisons are done simultaneously because if you ignore the analysis of variance results and run multiple comparisons, you will likely make type I error.

CHAPTER THREE

METHODOLOGY

3.0 ANOVA(Analysis Of Variance)

Given that $x_{\ell 1}, x_{\ell 2}, \ldots, x_{\ell n_\ell,}$ is a random sample from an N(μ_ℓ,σ^2) population,$\ell = 1,2, \ldots$, g, and that the random samples are independent.

Since populations usually corresponds to different sets of experimental conditions and therefore, it is convenient to investigate the deviations(τ_ℓ) associated with the ith population(production yield). In this case the decomposition become

$$\mu_\ell \quad = \quad \mu \quad + \quad \tau_\ell \tag{3.1}$$

The response(x_{ej}), distributed as $N(\mu + \tau_\ell, \sigma^2)$ can be expressed in the form

$$x_{ej} = \quad \mu \quad + \quad \tau_\ell \quad + \quad e_{ej} \quad (3.2)$$

$$\begin{pmatrix} overall \\ mean \end{pmatrix} \begin{pmatrix} treatment \\ effect \end{pmatrix} \begin{pmatrix} random \\ error \end{pmatrix}$$

Where e_{ej} are independent $N(0\,, \sigma^2)$ random variables.

To define uniquely the model parameters and their estimates, it is customary to impose the constraints

$\sum_{\ell=1}^{g} n_\ell \tau_\ell = 0$

Motivated by the decomposition in (3.2), the analysis of variance is based upon an analogous decomposition of the observations

$$x_{ej} = \quad \bar{x} \quad + \quad (\bar{x}_\ell - \bar{x}) \quad + \quad (x_{ej} - \bar{x}_\ell) \quad (3.3)$$

$$(observation) \begin{pmatrix} overall \\ mean \end{pmatrix} \begin{pmatrix} overall \\ mean \end{pmatrix} (residual)$$

Where $\bar{x}$ is an estimate of μ , $\bar{\tau}_\ell = (\bar{x}_\ell - \bar{x})$ is an estimate of τ_ℓ and $(x_{ej} - \bar{x}_\ell)$ is an estimate of e_{ej}.

The question of equality of means is answered by whether the contribution of the treatment array is large to the residuals. The size of an array is quantified by stringing the rows of the array out into a vector and calculating it's squared length. This quantity is called the sum of squares (SS).

From (3.3), the sum of squares satisfy the same decomposition as the observations.

$SS_{obs} \quad = \qquad SS_{mean} + \; SS_{treatment} + \qquad SS_{residuals}$

$\left(x_{ej} - \bar{x}\right)^2 = \; (\bar{x}_\ell - \bar{x})^2 \; + (x_{ej} - \bar{x}_\ell)^2 + 2(\bar{x}_\ell - \bar{x})(x_{ej} - \bar{x}_\ell)$

When we sum both sides over j, we get $\sum_{j=1}^{n_\ell}(x_{ej} - \bar{x}_\ell) = 0$

and obtain

$\sum_{j=1}^{n_\ell}(x_{ej} - \bar{x})^2 = n_\ell(\bar{x}_\ell - \bar{x})^2 + \; \sum_{j=1}^{n_\ell}(x_{ej} - \bar{x}_\ell)^2$ (3.4)

Next, summing both sides over ℓ we get

$\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(x_{ej} - \bar{x})^2 \; = \qquad \sum_{\ell=1}^{g} n_\ell(\bar{x}_\ell - \bar{x})^2 + \qquad \sum_{\ell=1}^{g}\left(x_{ej} - \bar{x}_\ell\right)^2$ (3.5)

(3.4)$(SS_{treatment})(SS_{treatment})(SS_{treatment})$

The above is summarize in ANOVA table by attribution of g−1 degree of freedom(df) to $SS_{treatment}$ and n−g= $(n_1 + n_2 + n_3 \ldots + n_g)$ – g degree of freedom to $SS_{residuals}$.

The total degrees of freedom is $n = n_1 + n_2 + n_3 \ldots + n_g$

Table 3.1 ANOVA TABLE FOR ONE WAY ANALYSIS

Source of variation	Sum of squares(SS)	Degrees of freedom(d.f)
Treatments	$SS_{tr} = \sum_{\ell=1}^{g} n_\ell(\bar{x}_\ell - \bar{x})^2$	$g - 1$
Residual (Error)	$SS_{res} = \sum_{\ell=1}^{g}\left(x_{ej} - \bar{x}_\ell\right)^2$	$\sum_{\ell=1}^{g} n_\ell - g$
Total	$SS_{total} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(x_{ej} - \bar{x})^2$	$\sum_{\ell=1}^{g} n_\ell - 1$

With the table 3.1 we can calculate F- statistics

$$F = \frac{SS_{tr}/_{(g-1)}}{SS_{res}/_{\sum_{\ell=1}^{g} n_\ell - g}} \qquad (3.6)$$

The Null hypothesis H_0 :τ_1 = τ_2 = …. τ_g = 0 at α level if

$F > F_{g-1\,,\,\sum_{\ell=1}^{g} n_\ell - g}at\ (\alpha)$

If Fin equation (3.6)is significant then the null hypothesis(H_o) is rejected, then there exist simultaneous difference between the means and hence we use the multiple comparison post hoc tests to check where the difference lies.

3.1.0 Factor Effects Modelfor A Single Factor (region) ANOVA

Given $y_{ij} = \mu_i + \epsilon_{ij}$

Wherey_{ij} is value of the response variable in the jth trial for the ith factor level.

μ_i is the unknown mean for all of the observations at level i.

ϵ_{ij} are independent normal errors with means 0 and variance σ^2

If we estimate all the cell means

μ_1, μ_2 , μ_r and also σ^2

The F-test answers the question of whether μ_i depends on i.

That is we test the null hypothesis $Ho: \mu_1 = \mu_2 = \cdots \mu$ against the alternative that not all the means are the same.

We re-parameterize cell means model by taking $\mu = \mu + \tau_i$

And hence $\tau_i = \mu - \mu i$= $1,2,\dots,r$

This means factor effect(τ_i) is the difference between the overall mean and the factor level mean.

Since most of the time we are interested in difference between means.

i.e $\mu - \mu i$= $1,2,\dots,r-1$

we consider the overparameterized model

$y_{ij} = \mu + \tau_i + \epsilon_{ij}$

where the usual assumptions apply, with $j = 1,2,\ldots n_i$ and $i = 1,2,\ldots,r$

If $\mu = \mu_r$ then $\tau_i = \mu - \mu_r$ $i = 1,2,\ldots,r-1$ and

Taking $\mu = \frac{(\mu_{1+} \mu_{2+} \ldots \mu_r)}{r}$

And hence we have $\tau_i = \mu - \mu$, $i = 1,2,\ldots,r$ and $\sum_{i=1}^{r} \tau_i = 0$

this is so because we have overparametrized model and in view of this we get extra parameters.

In this case one of the τ_i's become redundant.

To avoid redundancy and make the model equivalent we assume $\sum_{i=1}^{r} \tau_i = 0$

Thus $\sum_{i=1}^{r} \tau_i = 0$ becomes a convenient constraint because it means that they represent differences from the overall means.

The null hypothesis for F-test becomes $H_0 : \tau_1 = \tau_2 = \ldots = \tau_r = 0$

Since is an ($n_T \times (r+1)$)matrix andthe *rank(x)* =*r* this means the design is not full rank. This is a direct consequence that the model is overparametrized.

Since $X'X$ is not a full rank matrix we need a generalized inverse G

That is a matrix that satisfies

$$X'XGX'X = X'X$$

Such a generalized inverse is not unique. For each G we obtain a solution of the set of normal equations with $\beta = (\mu, \tau_1, \ldots . \tau_r)'$

And hence $X'X\beta = X'Y$

This solution takes the form $\beta = GX'Y$ (3.7)

3.2 MULTIPLE COMPARISON TEST (post Hoc Tests)

The analysis of variance (ANOVA) is a powerful procedure for testing the homogeneity of a set of means. However when the null hypothesis rejected in ANOVA and accept the stated alternative that the means are not equal, the ANOVA will not be able to identify which means are different.

With only two groups of observations we could compare the two groups using a t-test. When we have more than two groups, it is inappropriate to simply compare each pair using a t-test because of the problem of multiple testing. The appropriate procedure to do the analysis is to use a one-way analysis of variance (ANOVA) to evaluate whether there is any evidence that the means of the populations differ, we might then be interested in investigating which of the means are different. In this case we employ multiple comparison test, but when multiple comparison is done there is the likelihood of committing at least one type I error. In this case ANOVA and multiple comparisons are done simultaneously because if you ignore the analysis of variance results and run multiple comparisons, you will likely make type I error.

3.2.1 Experiment – wise error rate

The comparison of two means can be made through an F-test, a t-test or by the computation of confidence interval on the difference between the two means. However, serious difficulties occurs when the analysis attempts to make many or all possible paired Comparison.

For the case of K means, there will be, of course $\frac{K\,(k-1)}{2}$ possible paired comparison. Assuming independent comparisons the experiment – wise error rate (i.e., the probability of false rejection of at least one of the hypothers is given by 1 – $(1-\alpha)^r$, where α is the selected probability of type I error for a specific comparison, and $r = \frac{k\,(k-1)}{2}$. Clearly, this measure of experiment-wise type I error can be quite large. With the task of testing many paired comparison there is usually the need to make the effective contrast on a single comparison more conservative, multiple Comparison methods.

The first stage in the analysis is to find a significant *F* in the ANOVA. If *F* is not significant at a given level, post test which called for multiple comparison cannot continue and the analysis must end there. But if F is significant the post analysis can proceed.

There are range of multiple comparison test. Some of these ignore the problem of the type I error completely and others do not. Fisher LSD test takes no account of the number of comparisons being made and the increase risk of type I error is simply accepted. Other tests such as Newman-keuls and Duncan take account of the number of Comparisons being made and compute different values accordingly. At the more conservative end of scale, the Tukey and Scheffé tests allow all comparisons to be made as the test corrects for the increased risk of type I errors by reducing the significance level of the individual comparisons. The simplest and most conservative method is to apply a Bonferroni correction to the significant level. There are many multiple comparison tests but we shall consider the following;

i. The Tukey method (for all pair-wise comparisons)
ii. Fishers least significant difference (LSD)
iii. Scheffé method
iv. Bonfferoni's method

3.2.2 Tukey's Methods

Following an analysis of variance in which we have rejected the null hypothesis of equal means, the next step will be to test all pair-wise means;

H_o:$\mu_i = \mu_j$

H_o:$\mu_i \neq \mu_j i \neq j$

Tukey (1953) proposed a procedure for testing hypothesis for which the overall significant level is exactly α when the sample sizes are equal and at least α when sample sizes are unequal.

His procedure can also be used to construct confidence interval on the differences in all pairs of means. For these intervals, the simultaneous confidence levels is 100 (1 –α)% when sample sizes are equal and at least 100 (1 – α) % when sample size are not equal. Tukey's procedure makes use of the distribution of the studentized range statistics:

$$q = \frac{\bar{y}_{max} - \bar{y}_{min}}{\sqrt{\frac{MSE}{n}}} \quad (3.8)$$

where $\bar{y}_{max}$ and $\bar{y}_{min}$ are the largest and smallest sample means respectively out of a p sample means. MSE is the mean square error and n in the number of observation in a given group.

The studentized range statistic takes into account the number of samples. This is because as sample size (n) increases, the magnitude of the sample range also increases.

The studentized range statistic looks very similar to the t-statistics when there are more than two independent samples, when all simple sizes are equal and the assumptions of ANOVA hold the t-statistics in (3.8) becomes

$$t = \frac{\bar{y}_{max} - \bar{y}_{min}}{\sqrt{\frac{2MSE}{n}}}$$

$\therefore \qquad q = t\sqrt{2}$

Tukey's test declares two means significantly different if the absolute value of their sample difference exceeds the statistics:

$$T_\alpha = q_{\alpha(a,f)}\sqrt{\frac{MSE}{n}} \qquad (3.9)$$

Thus two means are significant if $|\bar{y}_i - \bar{y}_j| > T_\alpha$

Equivalently we could construct a set of 100 (1 –α) percent confidence intervals for all pairs of means as follows:

$$\bar{y}_i - \bar{y}_j - q_{\alpha(a,f)}\sqrt{\frac{MSE}{n}} < \mu_i - \mu_j < \bar{y}_i - \bar{y}_j + q_{\alpha(a,f)}\sqrt{\frac{MSE}{n}} \quad i \neq j \qquad (3.10)$$

When sample sizes are not equal (3.9) becomes

$$T_\alpha = q_{\alpha(a,f)}\sqrt{MSE\left(\frac{1}{ni} + \frac{1}{nj}\right)} \quad , i \neq j \qquad (3.11)$$

and the confidence interval becomes

$$\bar{y}_i - \bar{y}_j - \frac{q_{\alpha(a,f)}}{\sqrt{2}}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \leq \mu_i - \mu_j \leq \bar{y}_i - \bar{y}_j + \frac{q_{\alpha(a,f)}}{\sqrt{2}}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

(3.12)

Tukey's method is used when analyzing all possible combinations. Like both t-test and ANOVA, Tukey assumes that data from the different groups comes from population where the observations have a normal distribution and the standard deviation is the same for each group

3.2.3 FISHER'S LSD METHOD

After the rejection of the null hypothesis, the test does not indicate, which group differs. In order to analyze the pattern of difference between means the ANOVA is often followed by specific pair-wise comparisons. The first pair-wise comparison that was developed was by fisher in 1935 and is called the least significant difference (LSD) test. This technique can be used only if ANOVA is significant. The main idea of the LSD is to compute the smallest significant difference between two means as if these means had been the only means to be compared (i.e with a t-test) and to declare significant any mean difference larger than the LSD.

The rationale behind the LSD technique value comes from the observation that, when the null hypothesis is true, the value of the t-statistics evaluating, the difference between the groups y_i and y_j is equal to

$$t = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \qquad (3.13)$$

Where $\bar{y}_i =$ the mean of group y_i and $\bar{y}_j =$ the mean of y_j, MSE is the mean square error and n_i and n_j are the number of observations of ith and jth groups. It follows the student's t distribution with N-a degrees of freedom. The ratio t would therefore be declared significant at a given α level if the values of t is larger than the critical value for the α level obtained from the t distribution and $t_{v,\alpha}$ ($v = \mathbf{N} - a$).

Assuming a two sides alternative, the pair of means μ_i and μ_j would be declared significantly Different if

$$|\bar{y}_i - \bar{y}_j| > t_{\alpha/2, N-a}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \qquad (3.14)$$

the quantity in (3.15)

$$\text{LSD} = t_{\alpha/2, N-\text{a}}\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \qquad (3.15)$$

is called the least significant difference

if $n_1 = n_2 = n_3 \ldots . n_\text{a} = n$ then equation (3.15) now becomes

$$\text{LSD} = t_{\alpha/2, N-\text{a}}\sqrt{\frac{2MSE}{n}} \qquad (3.16)$$

To use the fisher LSD procedure, we compare the absolute value of difference between each pair of averages to the corresponding LSD.

If $|\bar{y}_i - \bar{y}_j| >$ LSD, it is concluded that the population means μ_i and μ_j differ at α-level (usually 0.5 or o.1).

LSD has more power compared to other comparison methods because the α level for each comparison is not corrected for multiple comparisons. And, because LSD does not correct for multiple comparisons. As a consequence, a revised version of the LSD has been proposed by Hayter (and is known as the Fisher-Hayter procedure) where the modified LSD (MLSD) is used instead of the LSD.

The MLSD is computed using the studentized range distribution q as

$$\text{MLSD} = q_{\alpha,\ \text{a}-1}\sqrt{\frac{MSE}{n}} \qquad (3.17)$$

The MLSD procedure is more conservative than the LSD, but more powerful than Tukey approach because the critical value for the Tukey approach is obtained from a studentized range distribution equal to “a” groups.This difference in range makes Tukey’s critical value always larger than the one used for MLSD and therefore it makes Tukey's approach more conservative

3.2.4 SHCEFFÉ METHOD FOR COMPAIRING ALL CONTRASTS

Scheffé procedure is perhaps the most popular of the post hoc procedures, the most flexible and the most conservative. This is because it correct α for all pair-wise or simple comparisons of means. Complex Comparisons involves contrasts of more than two means at a time. As a result Scheffé is also the least statistically powerful procedure. But Scheffé is a poor choice unless complex comparisons are made, because for simple or pair-wise comparisons scheffé will lead to Type II error. When all pairs of means are being compaired, Tukey's approach is the procedure of choice. In many exploratory experiments, the comparisons of interests are discovered only after preliminary examination of data. Scheffé(1953) has proposed a method for comparing any and all possible contrasts between treatment means.

In scheffé's method, the type I error is at most α for any of the possible comparisons. Scheffébased his argument using contrasts. In general contrast is a linear combination of parameters of the form:

$\Gamma = \sum_{i=1}^{a} c_i \mu_i$ thus the product of the contrast constants c_i and the treatment means μ_i. The contrast constants c_1, c_2... c_a sum to Zero. That is $\sum_{i=1}^{a} c_i = 0$. In this case, a hypothesis is expressed as

H_o : $\sum_{i=1}^{a} c_i \mu_i = 0$

H_i : $\sum_{i=1}^{a} c_i \mu_i \neq 0$

Testing hypothesis involving contrast can be done in two basic ways. The first method uses a *t*-test. In this method the contrasts of interest are written in terms of treatment totals, giving

$C = \sum_{i=1}^{a} c_i\, y_i$ the variance of *C* then becomes

$V(C) = n\sigma^2 \sum_{i=1}^{a} C_i^2$, when sample sizes of each treatment are equal.

If the null hypothesis above is true, the ratio

$$\frac{\sum_{i=1}^{a} c_i\, y_i}{\sqrt{n\sigma^2 \sum_{i=1}^{a} C_i^2}}$$

has the normal distribution with N(0, 1). If σ^2 is replaced by it's estimate, the *MSE* and use the statistic

$$t_o = \frac{\sum_{i=1}^{a} c_i y_i}{\sqrt{nMSE \sum_{i=1}^{a} C_i^2}} \quad (3.18)$$

to test the hypotheses above. The null hypothesis would be rejected if $|t_o|$ in (3.18) exceeds $t_{\alpha/2}, N - a$.

The second approach uses an F-test, with a degrees of freedom (1,*v*). The F statistic in this case can be written as

$$F = t_o^2 = \frac{\left(\sum_{i=1}^{a} c_i y_i\right)^2}{n\, MSE \sum_{i=1}^{a} C_i^2} \quad (3.19)$$

the null hypothesis is rejected if $F_o > F_{\alpha,1,N-a}$

Scheff$\acute{e}$ assumed that a set of *m* contrasts in the treatment means have been determined.

$\Gamma_u = c_1\mu_1 + c_2\mu_2 + \ldots + c_a\mu_a$, $\mu = 1, 2, 3 \ldots m$

The corresponding contrast in the treatment averages ($\overline{y}_i$) then becomes

$C_u = c_1 y_1 + c_2 y_2 + \ldots + c_a y_a$ and the standard error of this contrast is

$$S_{Cu} = \sqrt{MSE\sum_{i=1}^{a}\left(c_i^2/n_i\right)}$$

where n_i is the number of observations in the ith treatment. It can be shown that the critical value which C_u should be compared is

$$S_{\alpha,u} = S_{Cu}\sqrt{(a-1)F_{\alpha,a-1,N-a}} \qquad (3.20)$$

Therefore if $|C_u| > S_{\alpha,u}$, then the hypothesis that the contrast Γ_u equal zero is rejected. But the caution is that for all pairwise comparison of means then the best approach should be Tukey's approach.

3.2.4 BONFERRONI METHOD

The bonferroni method of paired comparisons allows any number of unplanned comparison between two means. It is based on ensuring that the probability of type one error across all tests is at least $1-\alpha$. In general if we have k independent significance at the α level, the probability p that we will get no significant difference in all these tests is simply the product of the individual probabilities:

$$(1-\alpha)^k$$

Thus if, with $\alpha = 0.05$, and there are $k = 10$ tests to conduct then we get

$p = (1-0.05)^{10} = 0.6$ tests of no significant difference in all individual probabilities.

This means we have about 40% chance that one of these 10 tests will turn out significant, despite each individual test only being at the overall significance is still at α level, Bonferroni adapt the significance level α^I of the individual tests. This results in the following relation between the overall and the individual significance level:

$(1-\alpha^I)^k = 1-\alpha$

Thus $\alpha^I = 1-(1-\alpha)^{1/k}$

and for small α reduce the equation to

$$\alpha^I = \frac{\alpha}{k}$$

Thus Bonferroni used $\frac{\alpha}{k}$ as a correction factor . While other multiple comparison procedured generally involve deriving a new test statistics to have adjustments, Bonferroni merely involves adjusting the critical value of the test statistic at hand example, the t-test or z-test and just replace α by $\frac{\alpha}{k}$. If

$H_o : \mu_i = \mu_j$

$H_1 : \mu_i \neq \mu_j i \neq j$

then if the t-test is use, the

$$t_{calculated} = \frac{\bar{y}_i - \bar{y}_j}{\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}} \quad (3.21)$$

and the $t_{critical} = t_{1-\alpha/2k}, N-a$ (3.22)

if $t_{calculated} > t_{critical}$, then the null hypothesis must be rejected indicating a significance difference between the means.

For $\frac{k(k-1)}{2}$ pairwise comparisons, the $1-\alpha$ pairwise Bonferronic Confidence interval (when using the *t*-test) for 100 (1 – α)% Confidence Interval on the *ith* treatment mean μ is

$$\bar{y}_i - t_{\frac{\alpha}{2k}}, N-a\sqrt{\frac{MSE}{n}} \leq \mu_i \leq \bar{y}_i + t_{\frac{\alpha}{2k}}, N-a\sqrt{\frac{MSE}{n}} \qquad (3.23)$$

A $100(1-\alpha)\%$ confidence interval on the difference in any two treatment means say $\mu_i - \mu_j$ would be

$$\bar{y}_i - \bar{y}_j - t_{\frac{\alpha}{2k}}, N-a\sqrt{\frac{2MSE}{n}} \leq \mu_i - \mu_j \leq \bar{y}_i - \bar{y}_j + t_{\frac{\alpha}{2k}}, N-a\sqrt{\frac{2MSE}{n}} \qquad (3.24)$$

Where *k* is the number of tests.

3.3 MIXED EFFECT MODEL

A mixed effect model is a statistical model containing both fixed effects and random effects.

It is denoted by $Y_i = X_i\beta + Z_ib_i + \mathcal{E}_i \quad i = 1,2,\ldots N$ (3.25)

Where Y_i is a vector of observations with mean $E(Y) = X\beta$

β is a vector of fixed effects.

b_i is a vector of independent and identically distributed (IID) random effects with mean E(b)=0 and variance-covariance matrix variable (b)=G.

$\mathcal{E}$ is a vector of IID random errors with mean E($\mathcal{E}$)=0 and variance var($\mathcal{E}$)=R.

X and Z are matrices of regressors relating the observations y to β and b

In matrix and vector notation $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$, $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$, $Z = \begin{bmatrix} z_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & z_N \end{bmatrix}$

$b = \begin{bmatrix} \mathrm{b}_1 \\ b_2 \\ \vdots \\ b_N \end{bmatrix}$ $and\ \mathcal{E} = \begin{bmatrix} \mathcal{E}_1 \\ \mathcal{E}_2 \\ \vdots \\ \mathcal{E}_N \end{bmatrix}$

In general, a linear mixed- effects model is any model which satisfies

$$\begin{cases} Y_i = X_i \beta + Z_i b_i + \mathcal{E}_i \\ b_i \sim N(0, D) \\ \mathcal{E}_i \sim N(0, \Sigma_{\mathrm{i}}) \\ b_i \dots b_N, \mathcal{E}_1 \dots . \mathcal{E}_N \ \ _{independent} \end{cases} \qquad (3.26)$$

Where Y_i is the n – dimensional response vectors for subject i, $1 \leq i \leq N$,*N* is the number of subjects, X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ dimensional matrices of known covariates, β is a *p*-dimensional vector containing the fixed effects. b_i is the *q*- dimensional vector containing the random effects and $\mathcal{E}_i$ is an n_i dimensional vector of residual components. Finally *D* is a general $(q \times q)$ covariance matrix with (*i,j*) element. $d_{ij} = d_{ji}$ and Σ_{i} is a ($n_i \times n_i$) covariance matrix which depends on *i* only through its dimension n_i i.e the set of unknown parameter in Σ_{i} will not depend upon *i*. it follows from (3.21) that, conditional on the random effect b_i , Y_i is normally distributed with mean vector $X_i \beta + Z_i b_i$ and with covariance matrix *D*. Further, b_i is assumed to be normally distributed with mean vector 0 and covariance matrix *D*.

3.4**Estimation of parameters in the mixed model**

The two most commonly used estimates are maximum likelihood estimation and restricted maximum likelihood estimation.

3.4.1 **Maximum likelihood estimation**

Let α denote the vector of all variance and covariance parameters. Consist of the $\frac{q(q+1)}{2}$ different element in D and of all parameters in Σ_i. Finally let θ =(β', α') be the s-dimensional vector of all parameters in the model for Y_i and let $\theta = \theta_\beta \times \theta_\alpha$ denote the parameter space for θ, with θ_β and θ_α the parameter spaces for the fixed effects and for the variance components respectively. The MLE approach is based on estimations obtained from maximizing the likelihood function

$$L(\theta) = \prod_{i=1}^{N} \left\{ 2\pi^{-n_i/2} \, |V_i(\alpha)|^{-1/2} \right.$$

$$\left. \times\, exp\left(-\frac{1}{2}(Y_i - X_i\,\beta)'\, V_i^{-1}(\alpha)(Y_i - X_i\,\beta)\right) \right. \qquad (3.27)$$

With respect to θ. The maximum likelihood estimator (MLE) of β is obtained from maximizing (3.27), conditional on α is then given by

$$\hat{\beta}(\alpha) = \left(\sum_{i=1}^{n} X_i{}'W_iX_i\right)^{-1} \sum_{i=1}^{n} X_i{}'W_iy_i$$

Where $W_i = V_i^{-1}$

3.4.2 Restricted maximum likelihood estimator (REML) for mixed model

Given $Y = X\beta + Zb + \varepsilon$

Where the vector Y, b and ε , and the matrix X are obtained from stacking the vectors Y_i, b_i and $\mathcal{E}_i$ the matrices X_i respectively underneath each other and where Z is the block- diagonal matrix with blocksZ_i on the main diagonal and zeros elsewhere. The dimension of Y equals $\sum_{i=1}^{n} n_i$. The maginal distribution for Y is normal with mean vector $X\beta$ and with covariance matrix $V(\alpha)$ equal to the block-diagonal matrix with blocks V_i on the main diagonal and zeros elsewhere.

REML estimation for the variance component is now obtained from maximizing the likelihood function of a set of error contrasts $V = A'Y$ where A is any $(n \times n - p)$ full rank matrix with columns orthogonal to the columns of the X matrix. It is based on the likelihood function

$$L(\alpha) = C|\sum_{i=1}^{n} X_i{}' W_i(\alpha) X_i|^{-1/2} L(\hat{\beta}(\alpha), \alpha)$$

where C is a constant not depending on $V_i^{-1}(\alpha)$ and where $L(\alpha) = L(\theta)$

3.5.0 Akaike information Criterion (AIC) for model selction

Akaike (1974) information criterion (AIC) is a very popular criterion for model selection among several models. It is based on the criterion that

$$\text{AIC} = -2l_{max} + 2k \tag{3.28}$$

reaches a maximum, where l_{max} is the log-likelihood maximum and k is the number of unknown parameters. The smaller the AIC, the better the model. AIC works poorly when in the case of multicolinearlity since when multicolinearlity exists between variables, the affected variables produces the same results of AIC.

3.6 RANDOM EFFECT MODELS

Random effect model represent a natural heterogeneity between subjects. The assumption of random effect model is justified for data where the between –subjects variability is large in comparison to the within-subject variability. Random effect models are used when there a lot of variation in data under study and linear models cannot be to analyze such data. Since the random effects in model (3.26) were assumed to random variables we used random effect model.

The distribution of vector $Y_i$of response for the ith observation, conditional on that observation's specific regression coefficientsb_i, is multivariate normal with mean vector $X_i + Z_i b_i$ and with covariance matrix Σ_i. The marginal distribution of b_i is normal with mean vector 0 and covariance matrix D. If we denote the density function of $Y_i$conditional on b_i and the prior density function of b_i by $f(y_i/b_i)$ and $f(b_i)$ respectively. The density function of b_i given $Y_i = y_i$ is given by

$$f(y_i/b_i) \equiv f(b_i/Y_i = y_i) = \frac{f(y_i/b_i)f(b_i)}{f(y_i/b_i)f(b_i)\,\mathrm{d}b_i} \qquad (3.29)$$

b_i is estimated by the mean of b_i. This estimate is given by

$$\hat{b}_i(\theta) = \mathrm{E}\left[b_i/Y_i = y_i\right]$$

$$= \int f(y_i/b_i)f(b_i)\,\mathrm{d}b_i$$

$$= D\acute{Z}_i W_i(\alpha)(y_i - X_i\beta) \qquad (3.30)$$

and the covariance matrix of the corresponding estimator equals

$$Var(\hat{b}_i(\theta)) = D\acute{Z}_i\left\{W_i - W_i X_i\left(\sum_{i=1}^{N}\acute{X}_i W_i X_i\right)^{-1}\acute{X}_i W_i\right\}Z_i D_i \qquad (3.31)$$

Where $W_i = V_i^{-1}$

But (3.31) underestimate the variability in $\hat{b}_i(\theta) - b_i$ since it ignores the variation of b_i. Therefore, inference for b_i is usually based on $Var(\hat{b}_i(\theta) - b_i) = D - Var(\hat{b}_i(\theta))$ as an estimator for the variation in $\hat{b}_i(\theta) - b_i$.

Random Intercept with Variance-Covariance Assumption (Autoregressive of Order1 (AR1))

Random intercept assumes that each region has different intercepts and the intercepts have an iid normal with mean Zero and some unknown variance. In using Autoregressive of order 1, which assumes all variances to be equal and all covariances decay or weakens exponentially as time increases. Since we are considering forty (40) years the matrix for this assumption becomes.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 \dots & \rho^{39} \\ \rho & 1 & \rho \dots & \rho^{38} \\ \rho^2 & \rho & 1 \dots & \rho^{37} \\ \vdots & \vdots & \vdots \ddots & \vdots \\ \rho^{39} & \rho^{38} & \rho^{37} \dots & 1 \end{bmatrix}$$

Random Intercept with Variance-Covariance Assumption (Compound Symmetry)

Random intercept of compound symmetry assumes constant correlation between regions of productions and the dependency is linearly correlated over time. The estimated variance in this assumption is zero (0) which suggests different regions have the same intercept. Furthermore, treating time as a random allows the covariance of the repeated measures to explicitly become functions of time. The matrix for this assumption is below; $\sigma^2 \begin{bmatrix} 1 & \rho & \rho \dots & \rho \\ \rho & 1 & \rho \dots & \rho \\ \rho & \rho & 1 \dots & \rho \\ \vdots & \vdots & \vdots \ddots & \rho \\ \rho & \rho & \rho \dots & 1 \end{bmatrix}$

1

CHAPTER FOUR

4.0 DATA PRESENTATION AND ANALYSIS

The research had two variables taking into consideration: the year of cocoa production and annual regional yield of cocoa produced in the six regions where cocoa has been produced in Ghana . The factor or treatment was the Six(6) regionswhere cocoa has been produced in Ghana. They are Ashanti, Brong-Ahafo, Central, Eastern, Volta and Western regions. Since the introduction of free mass spraying of cocoa in Ghana; Cocoa production has increased and the task is to indentify whether there is any significant difference in the mean cocoa production . Data collected from COCOBOD in Accra from 1969/1970 production year to 2010/11 production year was assessed and analysed using SAS system. The following were the results:

TABLE 4.1summary statistics on yield per region

	Mean	Minimum	Maximum	Std Dev
ASHANTI	88551.48	44928.00	145557.00	28084.25
BRONGAHAFO	55440.75	28756.00	119156.00	25914.14
CENTRAL	34872.25	13782.00	59713.00	14113.85
EASTERN	49109.75	25372.00	86000.00	18223.00
VOLTA	4897.00	906.0000	22188.00	5632.68
WESTERN	142823.85	31113.00	419710.00	119433.19

From table 4.1 it is shown that Western Region has the maximum mean of cocoa yield production over the forty (40) years of the research objective years.

It recorded a mean of 142823.85metric tonnes of cocoayield production with the maximum production of 419710.00 metric tonnes . The maximum yield production was realized in 2004/05 production year and in that year there was an increase in yield production for the other regions and this was reflected in table 4.3 at the appendix . In table 4.1, it was also shown that the minimum mean of cocoa yield production was recorded in Volta region with a value 4897 metric tonnes. Volta region recorded the minimum cocoa yield production among the six(6) regions with yield production of 906 metric tonnes as can seen in table 4.4 at the appendix. In the summary statistics table(table 4.1) above, the order of cocoa yield production among the six(6) regions over the 40 years in regard to quantity of yield produced over the years was in the order: Western, Ashanti, Brong Ahafo, Eastern, Central and Volta. The table 4.1 was also analysed to check the standard deviation in the means of yield produced among the six regions. It was seen that there were a lot of variations in the western region compared to the other regions. Fig 4.1 in appendix shows the residual plot for the variation in yield in the six regions.

4.1 Analysis of variance (ANOVA) on yield and year of production.

We used the null hypothesis to test that the means of Cocoa production yield in the six regions are all equal .

TABLE 4.2 ANOVA TABLE FOR YIELD OF COCOA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	4.57E11	9.15E10	33.72	<.0001
Error	234	6.35E11	2.71E10		
Corrected Total	239	1.09E12			

R-Square =0.418757

From table 4.2, it can be seen that the F = 33.72 with a p-value of 0.000l. From this P-value, since P< 0.01, we rejected the null hypothesis that all the six regions means are equal. We concluded that at least one of the group means is significantly different from the others .

The R-Square value of 0.418757 indicated that annual yield accounts for approximately 42% of the variance in the regions of production. We tested using post hoc multiple comparisons when the null hypothesis was rejected to check where the difference lies.

Table 4.3 in appendix also showed the model for the year of Cocoa production from 1969/70 to 2008/09 production years. It was seen that the year of production was not significant since it had a p- value of 0.8483 which is far greater than the critical value of 0.01 which was also confirmed in the R-square value of 0.12879 meaning the year of production account for only 12% of the production of cocoa in the six regions and hence the year of cocoa yield production has no or little influence on the annual yield of cocoa production among the six regions all other things being equal.

4.2.1Assessing Varieties for Assumptions

Independence of observation: The Independence of the observations of yields was ensured since the condition under which cocoa is produce does not depend on the other. The rainfall and sunshine pattern in one region is different from the other.

Test of Homogeneity : This assumption is a critical assumption since it checks for the equality of variance . We used the Levene's test to assess this assumption. Table 4.4 provides the Levene's test to check the assumption that the variances of the yields from the six(6) are equal, it was realized that the Levene test was significant; p value <.0001 .This is an indication that the assumption of homogeneity has been violated.Table 4.4 Levene's test for homogeneity of yield variance

Levene's Test for Homogeneity of yield Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Region	5	6.10E21	1.22E21	20.61	<.0001
Error	234	1.39E22	5.92E19		

Test For Normality

The shapiro-Wilk and Kolmogorov –Smirnov tests were used to test for normality . Both tests were significant which was a true indication of the violation from normality. The Fig 4.1 below shows the violation of this assumption from the normal probability plot.

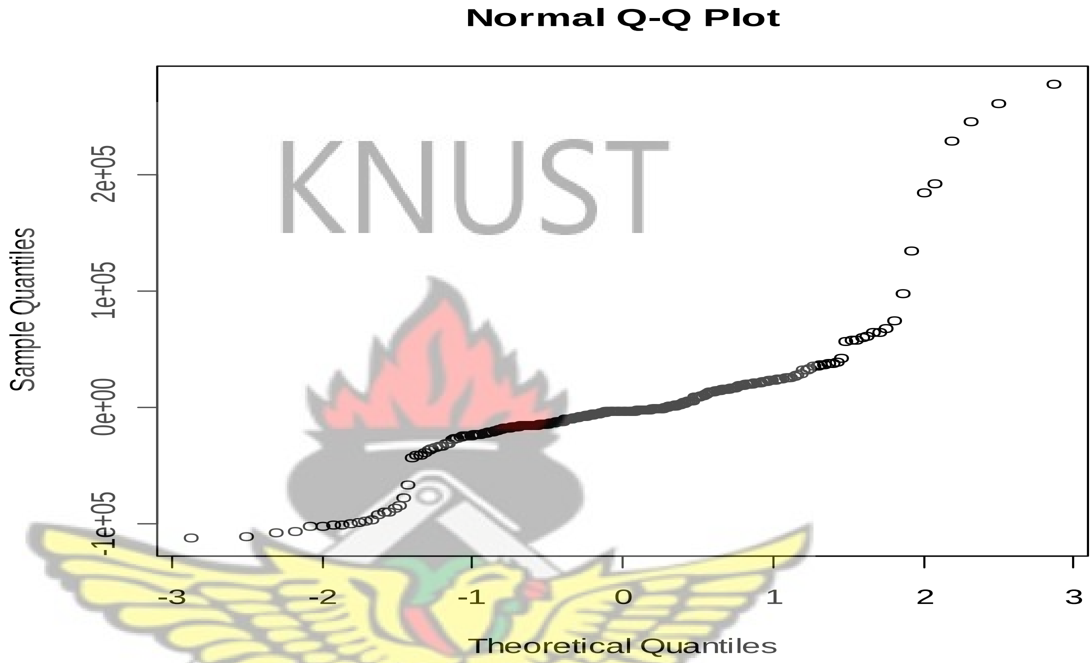


Fig 4.1 Normal probability plot

4.3 Transformation

Since the assumption of normality was violated we performed log transformation on yield of production and then tested for normality using the normal probability plot. After the transformation the deviation from normality was corrected. The figure 4.2 at the appendix showed the normal probability plot of the corrected log transformation.

4.4.0 Post Hoc Tests

Table 4.6 shows the results of the Post Hoc tests. Since the assumption of normality has been met, we looked at the multiple Comparisons(pairwise) tests using the three(3) tests :Least significant Difference(LSD),Bonferroni and Tukey. From the table 4.6 it is noticed that there is a degree of redundancy, so we are only concentrated on the unique pair Comparisons. All the three tests started with the highest mean of production (Western Region) and then compared with the next regions of mean of production. The table revealed that Western Region (mean = 142823.85) is significantly different from Ashanti Region (mean = 88551.48) at an alpha level of 0.05 significance with a mean difference of 54272 Metric tonnes of production.

Western Region is significantly different from Brong Ahafo with a mean difference of 87383 metric tonnes. Western Region is also significantly different from Central, Eastern and Volta Regions with mean difference of 93714, 107952 and 137927 metric tonnes respectively.

Ashanti Region is significantly different from Brong Ahafo with a mean difference of 33111 metric tonnes. This means that Ashanti Region produces more cocoa than Brong Ahafo.

Ashanti Region is also significantly differently from Eastern, Central and Volta with mean differences of 39,442, 53679 and 83654 metric tonnes respectively.

Brong Ahafo has a significant difference of 50544 metric tonnes from Volta Region.

Central Region was also significantly different from Volta Region by a value of 29975 metric tonnes. Easter Region was also significantly difference from Volta Region by a value of 44213 metric tonnes.Among the three(3) multiple comparism tests it was seen that LSD gave the best confidence interval since it had narrow or small intervals.

Table 4.6 summary of the multiple comparison tests;Comparisons significant at 0.05 level are indicated as ***

region Comparison	Difference Between Means	95% Confidence Limits						
		LSD		Tukey		Bonferroni		
6 – 1	54272	31319	77225	20795	87749	19722	88823	***
6 – 2	87383	64430	110336	53906	120860	52832	121934	***
6 – 4	93714	70761	116667	60237	127191	59163	128265	***
6 – 3	107952	84999	130905	74475	141429	73401	142502	***
6 – 5	137927	114974	160880	104450	171404	103376	172478	***
1 – 2	33111	10158	56064	-366	66588	-1440	67661	***
1 – 4	39442	16489	62395	5965	72919	4891	73992	***
1 – 3	53679	30726	76632	20202	87156	19129	88230	***
1 – 5	83654	60702	106607	50177	117132	49104	118205	***
2 – 5	50544	27591	73497	17067	84021	15993	85094	***
4 – 5	44213	21260	67166	10736	77690	9662	78763	***

4.5 CONTRASTS

Since western region hadthe maximum mean of production of cocoa among the six regions we established a contrast between western region and the other regions using Scheffe's method for comparing all contrasts.

contrast(Γ)= $5\mu_6 - \mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5$

We stated the null hypothesis that

H_O: $5\mu_6 = \mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5$ or

H_O: $5\mu_6 - \mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5 = 0$

The numerical value of the contrasts (C) becomes

C=$5\bar{y}_6 - \bar{y}_1 - \bar{y}_2 - \bar{y}_3 - \bar{y}_4 - \bar{y}_5$

C= 5(142823.85)$-$88551.48$-$55440.75$-$ 34872.25 $-$49109.75$-$ 4897

$\therefore C = -90047.38$

$|C|$= 90047.38

And the standard error was found as

$$S_C = \sqrt{MSE \sum_{i=1}^{6} \left(c_i^2 / n_i \right)}$$

$$S_C = \sqrt{2.7146E9 \left(30/_{40} \right)}$$

$$= 45121.50263$$

From S_C , the 5% critical value is $S_{0.05,1} = S_C \sqrt{(a-1) F_{0.05,a-1,N-a}}$

$$S_{0.05,1} = 45121.50263 \sqrt{5 \left(F_{0.05,5,234} \right)}$$

$$= 45121.50263 \sqrt{5(2.21)}$$

$$= 149{,}99.8247$$

Since $|C| < S_{0.05,1}$, we concluded that the contrast $\Gamma = 5\mu_6 - \mu_1 - \mu_2 - \mu_3 - \mu_4 - \mu_5$ equals zero; that is there is no strong evidence to conclude that mean cocoa production from western region differs from the means of cocoa production from the other five(5) regions.

4.5.1 Establishing contrast by geographical locations

We also checked if the cocoa production from the northern sector (Ashanti and Brong Ahafo) of the country differs from the southern sector (western, Eastern, Central and Volta). Thus establishing the contrasts:

$\Gamma = 2\,\mu_1 + 2\mu_2 - \mu_3 - \mu_4 - \mu_5 - \mu_6$

Hence the hypothesis;

$H_O\colon\; 2\,\mu_1 + 2\mu_2 - \mu_3 - \mu_4 - \mu_5 - \mu_6 = 0$

The numerical value of the contrasts (C) becomes

$C\; = 2\,\bar{y}_1 - 2\,\bar{y}_2 - \bar{y}_3 - \bar{y}_4 - \bar{y}_5 - \bar{y}_6$

$= (88551.48) - 2(55440.75) - 34872.25 - 49109.75 - 4897 - 142823.85$

$= 56281.61$

$|C| = 56281.61$

And the standard error was found as

$$S_C = \sqrt{MSE \sum_{i=1}^{6} \left({c_i^2}/{n_i} \right)}$$

$$S_C = \sqrt{2.7146E9\left({12}/{40}\right)}$$

$$= \sqrt{814380000}$$

$$= 28537.34$$

From S_C , the 5% critical value is $S_{0.05,1} = S_C\sqrt{(a-1)F_{0.05,a-1,N-a}}$

$S_{0.05,1}$=28537.34$\sqrt{5(F_{0.05,5,234})}$

=28537.34$\sqrt{5(2.21)}$

= 94862.53 Since$|C| < S_{0.05,1}$, we concluded that the contrastΓ= $2\mu_1 + 2\mu_2 - \mu_3 - \mu_4 - \mu_5 - \mu_6$ equals zero; that is there is no strong evidence to conclude that the mean cocoa production from the northern sector differ from the southern sector.

4.6 MIXED EFFECT MODEL

From fig 4.2 below, we observed that from the reference year 0 (1969/70) to year 19 (1987/88) cocoa production decreases slowly over the years for all the six(6) regions namely Ashanti, Brong Ahafo, Central, Eastern, Volta and western. Within the 19 years period, the cocoa production in Ashanti region was relatively higher. From the same fig 4.2, we also observed that from year 20 (1988/89) cocoa production in Brong Ahafo increased sharply and in the other four(4) regions cocoa production increased moderately, however in Volta region the production was low and stable.

From the individual(regional) profile in figure 4.2, cocoa production varies between the regions at the onset(year 1969/70). Over the forty(40) years period, cocoa production varied between and within regions as well. A statistical analysis considered to explain these variations was the Random effect model.

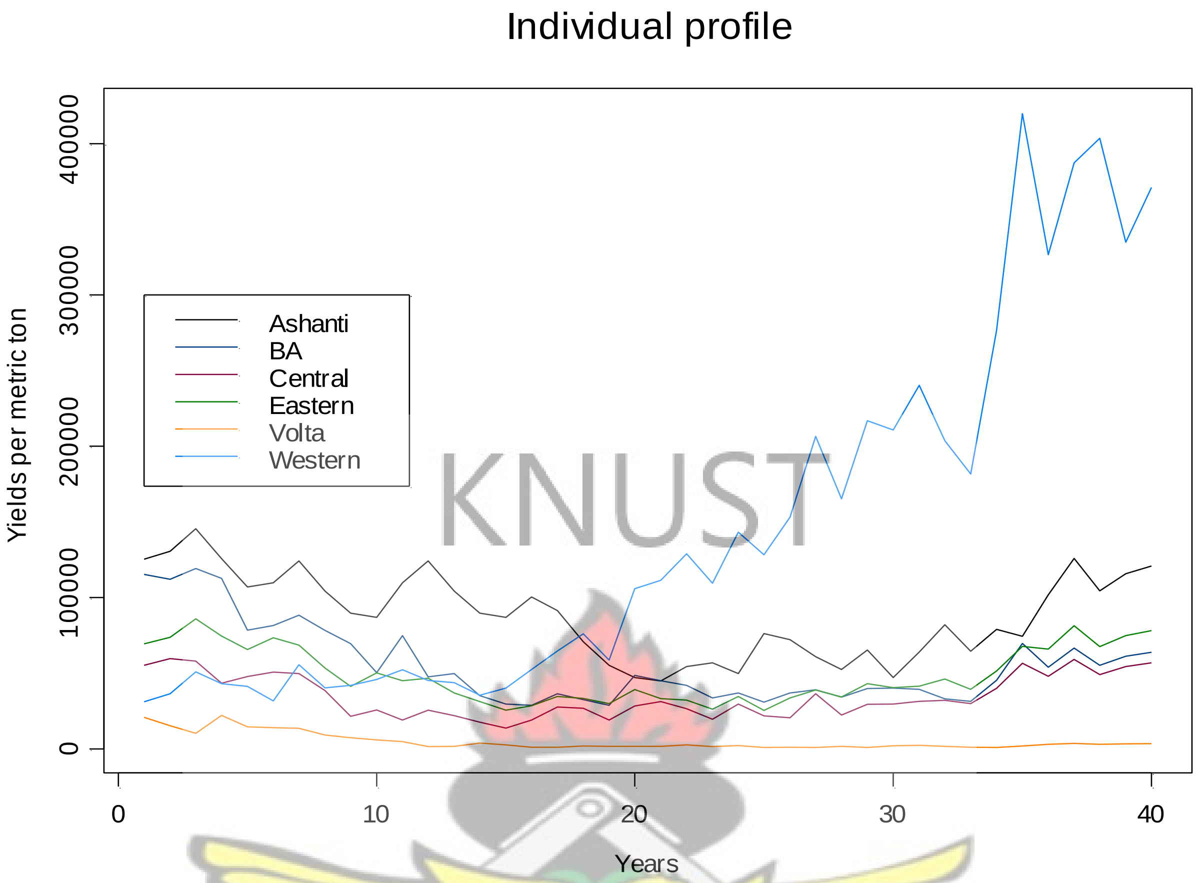


fig 4.2. Individual mean of cocoa production per region

Since a single regression line will not fit all of five regions yield of production, it made sense to use a random coefficient model. We restricted ourselves with,

Random intercept and random intercept and slopes models with linear and quadratictime effect .

RANDOM INTERCEPT WITH VARIANCE-COVARIANCE ASSUMPTION (AUT0REGRISSIVE OF ORDER 1) WITH LINEAR TIME EFFECT.

Random intercept assumed that each region has a different intercept which means at the base line year(1969/70 production year there were variations in cocoa production by regions . In this model we assumed that the intercepts have an iid normal with mean zero and some unknown variance. Table 4.7 at appendix showed that random intercept model has an AIC value of 5398.00 and here the year was not significant since the p-value was greater than 0.05. The estimated variance of the intercept is about 0.9766 which suggest that different regions have variations.

The model then becomes $Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + \varepsilon_{ij} \varepsilon_{ij} \sim N(0, \sigma^2)$, $bo_i \sim (0, V_{bo})$

$\hat{Y}_{ij}$ = $58294 + 1245.89 t_{ij} + b_{0i}$

i=1,…,6 regions j= 1,2,…,40 years

RANDOM INTERCEPT WITH VARIANCE-COVARIANCE ASSUMPTION (COMPOUND SYMMETRY) WITH LINEAR TIME EFFECT.

The random intercept assuming constant correlation between the six (6) regions and the dependency is linearly correlated over time. Here since there was constant correlation between the six regions, the estimated variance of the intercept is 0 which suggested different regions have the same intercept but the AIC value increased to 5870.7 and the year of cocoa production was significant which shows that this model is better than the first one with AR1. Table 4.8 at appendix indicate the AIC value.

The model then becomes $Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + \varepsilon_{ij} \varepsilon_{ij} \sim N(0, \sigma^2 \mathrm{I})$

$\hat{Y}_{ij}$ = $39816 + 1112.18 t_{ij} + b_{0i}$

i=1,…,6 regions j= 1,2,…,40 years

In comparing the two models above we noticed that neither the random intercept using AR1 with linear time effect nor the random intercept using compound symmetry with linear time effect model can completely explain variations in the data. We therefore considered a more complicated model that has both random intercept and slope.

RANDOM INTERCEPT AND SLOPE WITH VARIANCE – COVARIANCE ASSUMPTION (A RETROGRESSIVE OF ORDER 1) WITH LINEAR TIME EFFECT.

When both intercept and slope were random there was more flexibility in modeling the data because pairs of intercepts and slopes are assumed to have iid bivariance with normal distribution with mean zero and some unknown covariance matrix. The estimated variance of the intercept was about 0.9766

which is still the same as the random intercept model but the AIC value increased to 5400.00 and year was not significant. The model for this assumption becomes

$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij} \varepsilon_{ij} \sim N(0, \sigma^2)$

$\hat{Y}_{ij} = 58291 + 1245.88 t_{ij} + b_{0i} + b_{1i} t_{ij}$

i=1,…,6 regions j= 1,2,…,40 years

RANDOM INTERCEPT AND SLOPE VARIANCE-COVARIANCE ASSUMPTION (COMPOUND SYMMETRY) LINEAR TIME EFFECT.

The AIC value for this model was increased to 5872.7 and this is also at table 4.10 at appendix.

The model for this assumption also becomes

$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + \varepsilon_{ij} \varepsilon_{ij} \sim N(0, \sigma^2)$

$\hat{Y}_{ij} = 39816 + 1112.18t_{ij} + b_{0i}$

i=1,…,6 regions j= 1,2,…,40 years

RANDOM INTERCEPT WITH VARIANCE–COVARIANCE ASSUMPTION (AUTOREGRESSIVE OF ORDER I) WITH QUADRATIC TIME EFFECT.

Looking at the graph in figure 4.2, we recognized a curvelinear effect in the individual profiles as time increased, so we introduced a quadratic time effect in another model to check the behavior of the model. We noticed that the AIC value now reduced to 5382.7 and the quadratic time(year) was also significant.

The model for this assumption becomes $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t^2{}_{ij} + b_{0i} + \varepsilon_{ij} \varepsilon_{ij} \sim N(0, \sigma^2)$

$\hat{Y}_{ij} = 73629 - 3987.53t_{ij} + 127.63t^2{}_{ij} + b_{0i}$

i=1,…,6 regions j= 1,2,…,40 years

RANDOM INTERCEPT WITH VARIANCE- COVARIANCE (COMPOUND SYMMETRY) WITH QUADRATIC TIME EFFECT.

In this model also we recognized an AIC value of 5826.1 and the quadratic time was also significant. The model for the quadratic time effect is at appendix.

The model for this assumption becomes $Y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t^2{}_{ij} + b_{0i} + \varepsilon_{ij} \varepsilon_{ij} \sim N(0, \sigma^2)$

$\hat{Y}_{ij} = 85497 - 5413.66t_{ij} + 159.17t^2{}_{ij} + b_{0i}$

When we compared the AIC of the six random coefficient models we saw that the random intercept with variance – covariance assumption (AR1) with quadratic time effect model had the smallest AIC. It is therefore the best model we used to predict the response variable. The AIC values for the different models are presented in the table below.

Table 4.8 AIC values for the six(6) Random effectmodel Assumptions.

Model	AIC
1.Random intercept with variance – covariance assumption (AR1) with linear time effect.	5398.0
2. Random intercept with variance – covariance assumption (compound symmetry) with linear time effect	5870.7
3. Random intercept and slope with variance – covariance (AR1) with linear time effect	5400.00
4. Random intercept and slope with variance – covariance (compound symmetry) with linear time effect	5872.7
5.Random intercept with variance-covariance(AR1) with quadratic time effect	5382.7
6. Random intercept with variance-covariance(compound symmetry) with quadratic time effect	5826.0

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

This chapter presents the main findings of the study and concisely presents the conclusions that were made as a result of the study. The evidence presented by the data as well as the analysis given is also summarized in this section.

5.1 Discussion and Conclusion

The descriptive statistics used to analyze the results showed that majority (56.33%) of cocoa produced in the country comes from western region of Ghana with the mean annual production of 142823.9 metric tonnes while the lowest production(0.5%) of cocoa comes from volta region with mean annual production being 4897.00 metric tonnes.

The ANOVA analysis revealed that there is significant difference(significance= 0.0001) between the mean production of cocoa in the country and further analysis using the multiple comparison (pair wise tests) also showed that the mean annual production of cocoa differ in terms of the six region of cocoa production in the country.

Multiple comparison to detect where the difference lies using the three the comparison tests in the study(Tukey,LSD,Scheffe and Bonfferoni) showed that the mean difference between Ashanti and Western region was not significant but the mean difference between Ashanti and Central, Eastern, Volta ,Western were significant with mean difference of 53679.23,39441.73,83654.48 and -54272.38 respectively.

In comparing Brong Ahafo and the other regions it was seen that there was no significant difference between Ashanti,central and Eastern regions, the only significant difference was between Volta and Western regions with their mean difference values showed in table 4. 7

The significant difference between Central, Eastern,Volta and Western regions and the other regions are shown in table 4.7

The study also continued by establishing contrasts between first: Western region and the regions.

This revealed that the mean cocoa production from western region differs from the means of cocoa production from other five (5) regions. Secondly, establishing contrasts by geographical locations between the Northern sector (Ashanti and Brong Ahafo) and the Southern sector (western, Eastern, Central and Volta) revealed that there is no strong evidence to conclude that the mean cocoa production from the Northern sector differ from the Southern sector.

Further analysis using mixed effect model also revealed that from 1969/70 production, all the six regions with the exception of Volta region experienced increasing cocoa production trend as time increased. Western and Ashanti regions had the highest production over the years. Random intercept with variance-covariance assumption also showed that the different regions had variations in the mean cocoa production. Among the assumptions that were used in the mixed effect model, the model that best fit our analysis was random intercept with variance-covariance (AR1) with quadratic time effect. The AIC value for this assumption was 5382.7 which were the smallest among the six (6) assumptions. The quadratic time (year) was also significant. From this model as time increases the quantity of cocoa produced in the country also increases.

In conclusion , as time increases there exist variations in the production among the regions but these variations are constant and the variations weakens with time. Since variations are constant within regions but different between regions the policy by policy implementer for a specific year should be different for another year.

5.2 Recommendation

Based on the findings of the study, the following recommendations are made;

I. Investing in the cocoa bean industry in Volta region to raise its productivity especially among small holder farmers should be given the highest priority to increase revenue for both the government and the individual farmers.

II. Government should strive to make cocoa agrochemicals available at the right time in both Western and Ashanti region during the cocoa season and at subsidized prices. This would make it possible for the farmers to have access to input anytime they want to use it.

III. Cocoa diseases and pest control project (CODAPEC) should be strengthened to meet the recommended fungicides application per cocoa season to boost cocoa productivity in Volta region.

IV. There should also be improved extension linkage to sensitize cocoa farmers of the need to apply agrochemicals at the right proportion, recommended frequency per production season and at the right time. This will help to bridge the gap between potential and actual yield and hence, improve the level of efficiency and productivity.

V. There should be critical intervention by relevant stakeholders in the current production technology available to cocoa farmers in order to increase production to hit the 1.3 million metric tonnes by 2013.

REFERENCES

Anscombe, F. J., and J. W. Tukey (1963). "The examination and Analysis of Residuals." *Technometrics*, vol. 5, pp. 141-160.

Bancroft, T. A. (1968). *Topics in Intermediate Statistical Methods*. Iowa State University Press, Ames, Iowa.

Barnett, V., T. Lewis (1994). *Outliers in Statistical Data*, 3rd edition. Wiley, New Yorke.

Bennett, C. A., and N. L. Franklin (1954). *Statistical Analysis in Chemistry and the Chemical Industr*. Wiley, New York.

Bowker, A. H., and G. J. Lieberman (1972). *Engineering Statistics*. 2nd edition. Prentice-Hall, Englewood Cliffs, N.J.

Box, G. E. P. (1999). " Statistics as a Catalyst to Learning by Scientific Method Part 11- A Discussion" (with Discussion). *Journal of Quality technology*, Vol. 31, pp. 16-29.

Box, G. E. P., and D. R. Cox (1964). "An Analysis of Transformations." *Journal of the Royal statistical Society*, B, Vol.26, pp.211-243.

Box, G. E. P., and R. D. Meyer (1986). "An analysis of Unreplicated Fractional Factorials." *Technometrics*, Vol. 28, pp. 11-18.

Cochran, W. G. (1957). "Analysis of Covariance: Its Nature and Uses." *Biometrics*, Vol. 13, No. 3, pp. 261-281.

Cook, D. R. (1979). "Influencial Observations in Linear Regression."*Journal of the American Statistical Association*, Vol. 74, pp. 169-174.

Cornell, J. A. (1990). *Experiments and Mixtures: Designs, Models, and the Analysis of Mixture Data*. 2nd edition. Wiley, New York.

Duncan, A. J. (1986). *Quality Control and Industrial Statistics.* 5th edition . Richard D. Irwin, Homewood, I11.

Dunnett, C. W. (1964), " New Tables for Multiple comparisons with a Control", biometrics Vol. 20, pp 482- 491.

Eisenhart, C. (1947). " The Assumptions Underlying The Analysis Of Variance." *Biometrics* Vol. 3, pp. 1-21.

Fisher, R. A., and F. Yates (1953). *Statistical Tables for Biological, Agricultural, and Medical Research*. 4th edition. Oliver and Boyd, Edinburgh.

Levene, H. (1960), "Robust Tests for Equality of Variance." *In contributions to probability and Statistics.* Z. Olkin, ed. Stanford University Press, Palo Alto, CA. pp. 278-292.

McCullagh, P., and J.A. Nelder (1989). *Generalized Linear Models*. 2nd edition. Chapman and Hall, New York.

Montgomery, D.C., and G. C. Runger (1999). *Applied Statistics and Probability for Engineers.* 2nd edition. Wiley, New York.

Scheffe, H. (1956a). "A 'Mixed Model' for the Analysis of Variance." *Annals of Mathematical Statistics,* Vol. 27, pp. 23-36.

Scheffe, H. (1956b). "Alternative Models for the Analysis of Variance." *Annals of Mathematical Statistics,* Vol. 27, pp. 251-271.

Scheffe, H. (1959). *The Analysis of Variance.* Wiley, New York.

Searle, S. R., F. M. Speed, and H. V. Henderson (1981). "Some Computational and Model Equivalences in Analyses of Variance of Unequal- Subclass-Numbers Data." *The American statistician,* Vol. 35, pp16-33.

Abdulai, A. and Reider, P. (1995) "The Impact of Agricultural Policy on Cocoa Supply in Ghana: error- Correction Estimation. *Journal of African Economies.***4**:315-335

Abekoe, M. K., Obeng- Ofori, D. & Egyir, I. S. (2002). Technography Of Cocoa In The Forest Zone Of Ghana. Report Presented at the *Convergence Of Sciences' International Workshop,* 23-29 March 2002, Benin. Unpublished Project Document, 29 pp.

Adegbola, M. O. K. (1972). "Cocoa Diseases of West Africa" 7th international *Cocoa Research Conference,*Douala, United Republic of Cameroon.

Akayama, T and Duncan, R. C. (1982) Analysis of the World Cocoa Market," *World Bank Staff Commodity Working Paper*No**. 8** (Washingtton: World Bank)

Ali, F. M. (1969). Effects of rainfall on yield of cocoa in Ghana. *Experimental Agricultural,***5**: 209-213

Amos, T. T. (2007). An Analysis of Productivity And Technical Efficiency Of Smallholder Cocoa Farmers In Nigeria, *J. Soc. Sci*., **15**(2): 127-133 (2007)

Appiah, M. R, Ofori- Frimpong, K and Afrifa, A. A. (2001). Cocoa variety + fertilizer trial (k6-02). Annual report of the cocoa Research Institute of Ghana.

Appiah , M. R., Sackey, S. T., Ofori Frimpong, K and Afrifa, A. A. (1997). The consequences of Cocoa production on soil fertility in Ghana: a review, *Ghana Journal of Agricultural Science***30**: 183-190.

Bateman, M. J. (1990). Ghana cocoa pricing policy study. "*World bank working paper WPS429".*

Bateman, M. J. (1974) "An econometric Analysis of Ghanaian Cocoa Supply," In The Economics Of Cocoa Production And Marketing, ed. By R. A. Kotey, C. Okali and B. E. Rouke (Legon, Ghana: ISSER)

Brew, K. M. (1991). Relationship between yield, rainfall and total Sunshines hours. *Rep. Cocoa Research Institute Of Ghana.* 1988/89, **P30**-32

Chirwa, E. W., (2007), Sources of Technical Efficiency Among Smallholder Maize Farmers In SouthernMalawi, *African Economic Research consortium* (AERC) Research paper **172**

Coelli, T. J. (1996). A guide to FRONTIER Version 4.1: a computer Programme for Frontier Production Function Estimation. *CEPA Working Paper 96/07*. School of economics, university of New England, Armidale

COCOBOD (1998). Socio- Economic Study. Final Report. Cocoa Board, Ghana.

COCOBOD (2000) *Ghana Cocoa Board Handbook.* 8th ed. Jamieson's Cambridge Faxbooks Ltd, Accra. 62pp.

COCOBOD (2002). Socio- Economic Study. Final report. Cocoa Board, Ghana.

COCOBOD (2006). Socio- Economic Study. Final report. Cocoa Board, Ghana.

Gockowski, J. (2007) 'Cocoa Production Strategies and the Conservation of Globally Significant Rainforest Remnants in Ghana'. STCP-IITA mimeo.

Greene, W. H. (1980). Maximum Likelihood Estimation of Econometric Frontier Functions, *Journal of Econometrics*. 13:27-56

ISSER (2000). State of the Ghanaian Economy. Institute of Statistical, Social and Economic Research (ISSER), University of Ghana, Legon, Accra, Ghana.

Manu, J. E. A. (1974), "Cocoa in the Ghanaian Economy". In, R.A. Kotey, Okali and B.E Rouke (eds). Economics of Cocoa Production and Marketting. ISSER, University of Ghana, Legon, Accra, Ghana.

Ministry of Finance (1999). Ghana cocoa sector Development Strategy. Ministry of Finance, Accra.

Nkamleu, G. B. and Kielland, A. (2006). Modelling Farmers' Decisions on Child Labour and Schooling in the Cocoa Sector: A Multinomial Logit Analysis in Cote d'Ivoire. *Agricultural Economics,***35**:319-333.

Patton, M. Q. (1980): "*Qualitative Evaluation Methods*". Great Britain: Sage Pub. Inc.

APPENDIX

TABLES

TABLE 4.3 ANOVA MODEL FOR YEAR OF PRODUCTION

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	39	1.4062834E11	3.60585499E10	0.76	0.8483
Error	200	9.5222975E11	4.76114876E10		
Corrected Total	239	1.0928581E12			

FIGURES

FIG 4.1 Residuals Against Regions of yield production

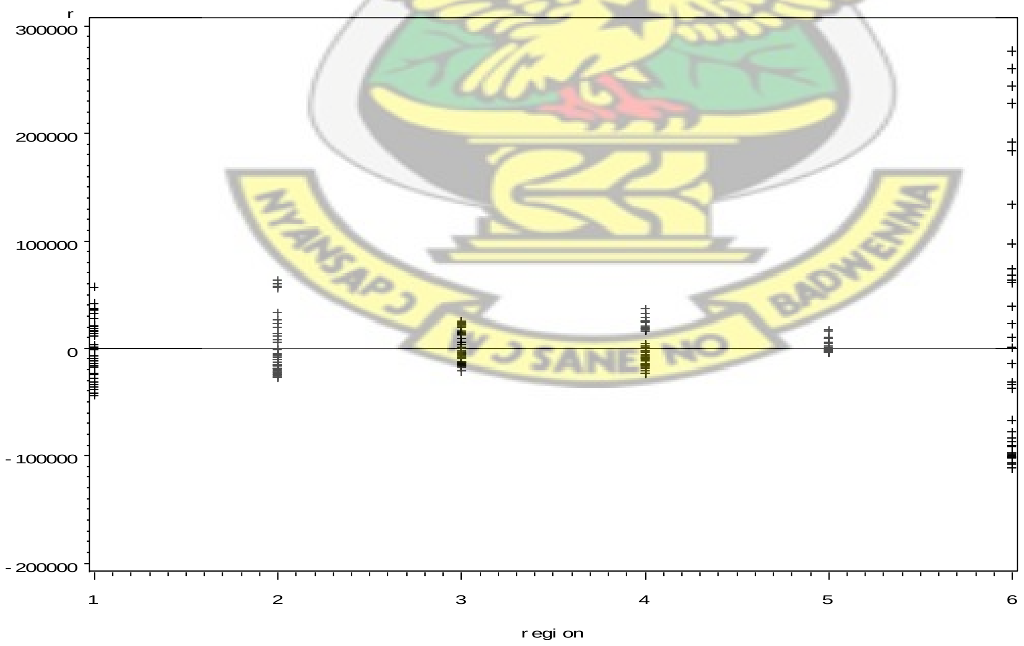
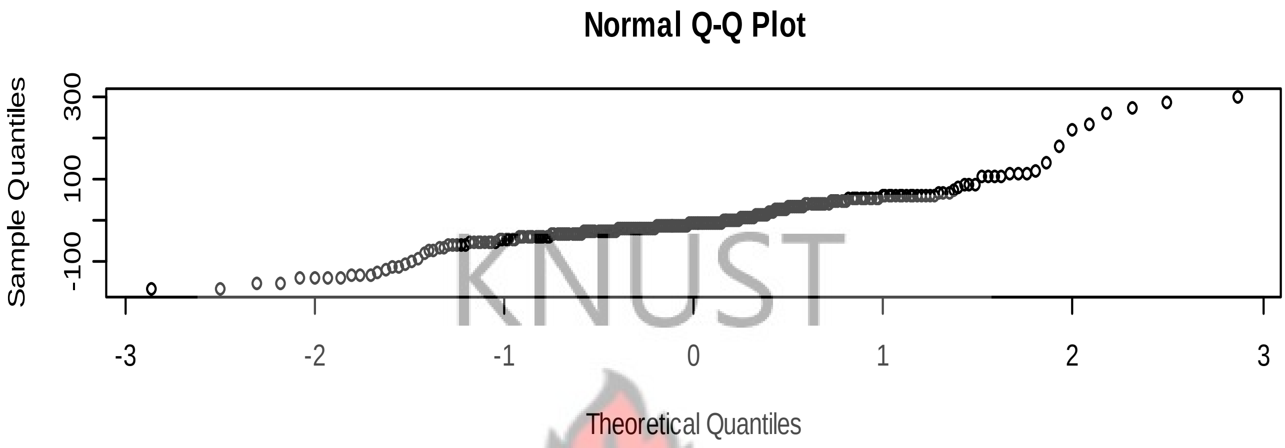
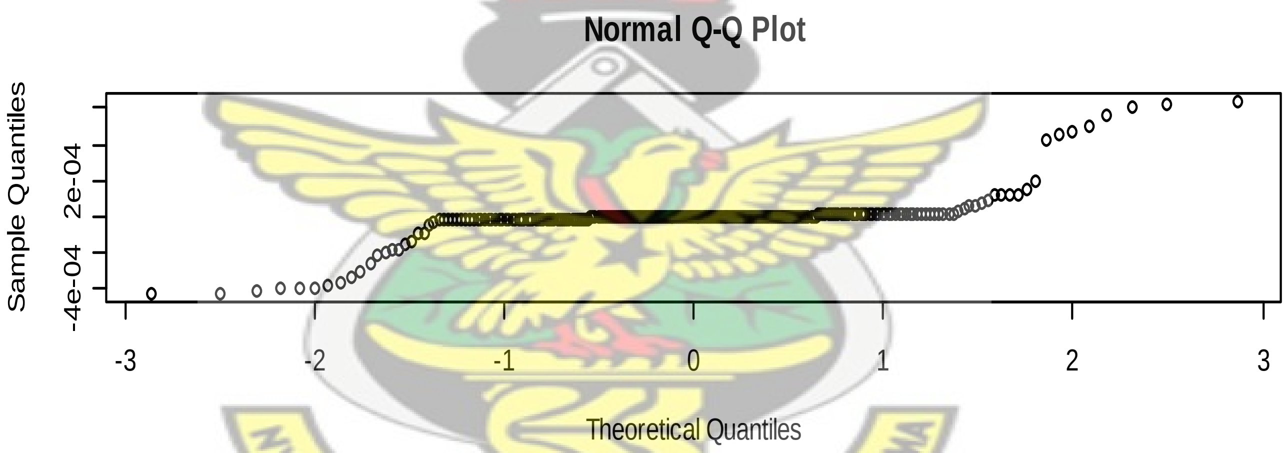
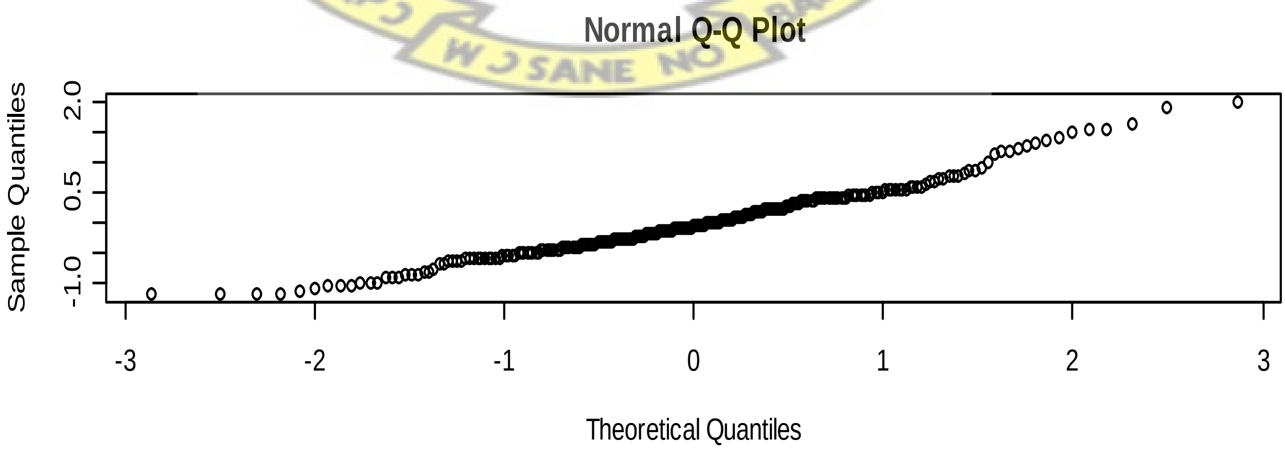


FIG 4.3 LOG TRANSFORMATION







Summary analysis on year of cocoa yield production

YEAR OF COCOA PRODUCTION	**Mean**	**Minimum**	**Maximum**	**Std Dev**
1	69576.17	20878.00	125406.00	43064.46
2	71315.67	15340.00	130544.00	43954.68
3	78310.67	10289.00	145557.00	49075.41
4	70307.17	22188.00	125648.00	41607.44
5	59145.17	14489.00	107028.00	32084.97
6	60213.83	14009.00	109802.00	34983.19
7	66720.17	13622.00	124315.00	37475.26
8	54018.50	9228.00	104215.00	33296.97
9	45223.17	7368.00	89619.00	30271.12
10	44179.33	5980.00	86913.00	27220.14
11	50976.50	4776.00	109802.00	38011.19
12	48458.67	1496.00	124315.00	41228.45
13	43051.17	1683.00	104215.00	34571.12
14	35489.17	3776.00	89619.00	29235.87
15	33129.00	2656.00	86913.00	29380.75
16	**38373.83**	**1028.00**	**100362.00**	**34636.02**
17	42685.50	1117.00	91537.00	31402.85
18	40280.83	1903.00	76037.00	28165.86
19	32286.83	1806.00	58742.00	21680.11
20	45154.67	1676.00	105894.00	34383.46
21	44642.50	1785.00	111513.00	36381.24
22	47810.33	2645.00	128955.00	43353.09
23	41233.33	1595.00	109469.00	38013.01
24	49422.33	2272.00	143274.00	48575.52
25	47291.50	923.000	128323.00	46790.90
26	52925.17	1067.00	153161.00	54367.92
27	63808.00	906.000	206585.00	72570.13
28	51736.83	1678.00	165361.00	58117.51
29	65970.67	976.000	216967.00	76885.71
30	61733.17	2062.00	210710.00	74706.87
31	69817.33	2352.00	240331.00	85872.11
32	66460.00	1680.00	203627.00	72044.54
33	58014.17	1079.00	181658.00	63890.14
34	82219.00	913.00	276586.00	98471.44
35	115031.67	1909.00	419710.00	151637.59
36	99886.33	2996.00	326628.00	115556.08
37	123409.67	3703.00	403550.00	142769.61
38	102422.00	3073.00	334919.00	118488.96
39	113463.50	3404.00	371026.00	131263.28
40	118440.33	3554.00	387299.00	137020.18

Tests for Normality

Test	Statistic		p Value	
Shapiro-Wilk	W	0.760469	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.181435	Pr > D	<0.0001