STATISTICAL ANALYSIS FOR A LATTICE DESIGN EXPERIMENT TO DETERMINE THE BEST PERFORMING MAIZE VARIETY IN MACHAKOS COUNTY

BY

ALFRED NJOROGE MAINA

I56/64926/2010

UNIVERSITY OF NAIROBI

COLLEGE OF BIOLOGICAL AND PHYSICAL SCIENCES

SCHOOL OF MATHEMATICS

THE DISSERTATION IS SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTER OF SCIENCE IN BIOMETRY

NOVEMBER 2013

DECLARATION

This project is my original work and has not been presented for a degree in any other University.

ALFRED NJOROGE MAINA

---------- ----------
Signature Date

DECLARATION BY SUPERVISORS

This project has been submitted for examination with our approval as supervisors

Mr. J.N.Mwangi
Kenya Agricultural
Research institute (KARI)
P.O.Box 57811 ---------- ----------
Nairobi. Kenya Signature Date

Prof. M.M.Manene
University of Nairobi
School of Mathematics
P.O.Box 30197 ---------- ----------
Nairobi.Kenya Signature Date

DEDICATION

I dedicate this master degree to my wife, daughters and mum for their outstanding support, prayers and providing conducive environment during my study period.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all the People and University of Nairobi that have contributed to the success of my master`s degree in Biometry. May I thank the teaching and non-teaching staff in the school of Mathematics, University of Nairobi for their support. I would like to recognize the contribution of my Msc. lecturers : Mr. Mwangi, Mr. Nderitu, Mrs wang`ombe, Mrs Obudho, Dr Kipchirchir, Dr Owour, Dr Oeba, Prof Weke, and Prof. J.A.M Otieno and Prof .M.Manene.

In particular, I appreciate the extra tireless effort made by my supervisors Mr.J.N. Mwangi and Prof.M.Manene for their guidance and support right from the start and throughout of my project formulation. I also appreciate Mr. Mwangi for his facilitation in the acquisition of the data from KARI that I have used in my study.

I appreciate the moral support and sacrifice shouldered by my wife, Janine and my daughters, Sasha and Candy who comprehended my occasional absence in most of the family matters. A lot of encouragement came from my Mum, sisters and my In-laws: Rev.Murimi and Mrs. Murimi. I would like to thank my classmates Otieno Otieno, Mulinge (K.C.A), Joyce (Wrigley), Karanja (Strathmore), Waweru (K.I.E) and Mbugua (St Georges) for their friendly and encouragement as well as suggestions offered to me throughout the course. Finally I would like to thank the following from KIST: Koigi, Lillian, Kageni ,Wanjau and Joe (K.U) for assisting me partially in the preparation of this script.

Above all, I thank the Almighty God for bestowing on me good health throughout the study period.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDICES

ABBREVIATIONS AND ACRONYMS

ANOVA	Analysis of variance
ASAL	Arid and Semi Arid Land
CIMMYT	International Maize and Wheat Improvement Centre
CRD	Completely Randomised Design
df	Degree of freedom
EDA	Exploratory Data Analysis
FAO	Food and Agriculture Organisation
GDP	Gross Domestic Product
ha	Hectare
H_1	Alternative Hypothesis
KARI	Kenya Agricultural Research Institute
Kg	Kilogram
LSD	Least Significant Difference
MS	Mean Square
MSE	Mean Square Error
Q-Q	Quantile – Quantile plots
RCBD	Randomised Complete Block Design
SAS	Statistical Analysis System
SPSS	Statistical Package for Social Scientists
SS	Sum of Squares
ton	tonne
UNESCO	United Nation Educational, Scientific and Cultural Organisation

ABSTRACT

Maize is one of Kenya's main staple food crop which plays a major role in the livelihoods of many households in terms of food security, income and employment generation. Over eighty percent of Kenya's population which is currently over forty million depends on availability of maize as the main food. There is a shortage of maize as consumption outstrips production, which is caused by low maize production, especially in semi arid or marginal regions. Feeding the growing population and meeting the increased domestic future demand will continue to be a challenge unless maize production per hectare is improved. This can be achieved through application of new farming technologies that could enhance maize production such as; planting the best performing and improved hybrid maize varieties which are in the market. Most of small scale farmers do not know nor plant these hybrid maize varieties which give optimal yield. This study, therefore seeks to determine the best performing maize variety out of the twenty five selected. The field trials were conducted in two locations namely Katumani and Kangundo in Machakos County, Kenya. The broad objective of the study was to determine the overall best maize variety across the two stated locations. The design used was a partially balanced lattice design carried out in two locations, each having two replications. Results revealed that there were significant mean differences among the maize varieties at 5% significant level in Katumani. No significant mean differences were noted in Kangundo. In Katumani, the best variety was T9 with mean yield of 1.158 t/ha while in Kangundo, T16 with mean yield of 1.747 t/ha. Varieties T16, T2, T5 and T14 were among the top ten in both locations. Farmers in Kangundo should be encouraged to plant varieties T16 and T2 while T9 and T22 in Katumani. Most of the varieties do not differ very much in yield capacity and more research should be conducted, based on the diversity of the farmer's requirements.

Keywords: Maize variety, yields, partially balanced lattice.

CHAPTER 1

1.0 INTRODUCTION

Maize is the primary staple food in Kenya. It is also the key food crop, accounting for 2.4% of Kenya's gross domestic product (GDP) and 12.6% of agricultural GDP (De Groote et al, 2005). Over eighty percent of Kenya's population depend on maize production for food. Demand for maize continue to rise and that is why prices have soared by 25% above 2011 (Eriyo, 2013). Kenya's poorest citizens spend nearly a third of their income on maize and therefore improving maize production is considered to be one of the most important strategy for addressing food insecurity problem in a country. However, despite the effort made by Kenya Agricultural Research Institute (KARI) that has led to the development and release of several high yields maize varieties, their adoption by farmers has been low. This has been attributed to lack of sufficient information or exposure to the varieties (Mureithi, 2005). This has led to low production especially in marginal or semi arid areas. There is a scarcity of land for cultivation due to high population density. The best mechanism of increasing maize production in future is by improving yield per hectare on land under use. The average maize yield is about 1.8 t/ha but yields of over 6 t/ha can be achieved (Makokha et al, 2010). This yield potential can be achieved by small or large scale farmers adopting and planting maize varieties recommended for their areas by research bodies like KARI. This study seek to investigate through field trials the best yield maize variety, if planted by farmers in Machakos county could lead to increase in maize yield and boast food security.

1.1 Background Information

Maize is known by its scientific name as *'zea mays'* and is not an indigenous cereal in Kenya (NAFIS, Kenya). It originated in Central America about 6,000 to 7,000 years ago. Maize arrived in Africa most likely through Portuguese traders who stopped along the African Coast

during 16th century (Miracle 1965). This maize was flinty, low yielding and varied in colour. From the coast, maize slowly moved inward particularly through the routes of slave traders. Maize became an important food crop in East Africa at the beginning of 20th century, when European settlers introduced new white dent varieties imported from South Africa. By the 1930s, maize was a dominant food crop; its expansion was driven by the demand for starch industry in England and the need to feed miners and farm workers (Byrelee and Eicher, 1997; Snake and Jayne, 2003). The present maize varieties in the country are mainly the result of maize breeding and agronomic research programmes which began intermittently in the 1930s. From 1990 to 2010, maize area in Kenya has increased from half a million hectares to about 1.6 million hectares. The total production increased from 0.8 million tonnes to about 2.5 million tonnes in the same period. However, average yields decreased slightly from 1.7 tons/ha in 1990 to 1.6 tons/ha in 2010 (FAO, 2011). In a typical year, maize provides 42% of dietary energy intake for Kenyan consumers, including both rural and urban areas (Mohammed and Underwood, 2004). Apart from food for humans, maize is used for many different purposes including feed for livestock, and raw material for agro-allied industries. Maize is eaten in form of grains and processed to offer various product ranges, which include maize flour that is used to make Kenya's common meal "ugali" and porridge. Maize is also used to make vegetable oils and sometimes fermented to produce alcohol to make local beer "busaa''. Green maize, fresh on the cob, is eaten roasted or boiled separately or mixed with legumes. Maize remnants after harvesting are used as fodder and can also be used to make silage when completely dried. It is therefore very important to invest in maize due to its varied and wide economical importance. Every part of maize plant has economical value; the grain, leaves, stalk and cob can be used to produce a large variety of food and non-food products.

1.1.1 Ecological requirements

Maize is grown at latitude ranging from the equator to approximately 50^{o} North and South. It can grow in a wide range of agro-ecological zones in Kenya ranging from 0-2200m above sea level depending on variety. This reflects its ability to adapt to a wide range of production environments, under temperature ranging from extreme cold to very hot, under moisture regimes ranging from extremely wet to semi-arid. Cold conditions extend the maturity period whereas high temperatures lower the yields. The optimum temperature for good yield is 30^{o}. Very low or high altitudes results in poor yields. Maize can grow on a wide range of soils though it performs best in well drained and well aerated loam or silty loam or alluvial soils with PH of 5.5-7. It is intolerant of water logging. Maize is grown on terrain ranging from completely flat to precipitously steep, in many different types of soil (Morris, 1998b). Maize grows well with 600-900mm of rainfall but average rainfall range is between 250mm to 2100mm per season (NAFIS, Kenya). The rainfall should be well distributed throughout the growing period. The rainfall is most critical at flowering and silking stage. The wide range of conditions has led to a continuous interaction of genotype with environment and formation of new maize types in farmer's field both through natural crossing and farmer selections. The performance of maize varieties is therefore highly specific to each condition (Smale et al 2011).

1.1.2 Agronomic practices in maize variety trials

Good management practices are essential for the production of high yields in maize variety trials. The management practices include seed dressing, thinning, the filling of vacancies in plant stands, cultivation, control of weeds, diseases, insects and vertebrate pests, fertilizers application and timely harvesting.

1.2 Literature Review

A correct experimental design is as important as a correct statistical analysis in order to obtain valid and reliable conclusions from trials or field experiments. Certain restrictions must be imposed when plots are arranged in order to be able to estimate the errors accurately. The choice of experimental design as well as of statistical analysis is of huge importance. These are necessary in order to obtain precise results (Mohsen and Hegazy, 2013).

The primary aim of most agricultural field experiments is the efficient estimation of treatment effects. To achieve this, it is important to control field variation that is caused by experimental management, fertility trends and other environmental factors.

Fisher (1926) in his first paper in field experimental designs emphasized the importance of randomized arrangements in the estimation of experimental error and described the Randomized complete block design (RCBD) and Latin square design as adequate. However, in some situations efficiency of the randomized block, Latin square and other complete block types of experiment is not high.

The problem with complete blocking is that as the block size increases due to the increase in the number of treatments, the homogeneity of experimental plots, within a large block is difficult to maintain and thus local control of experimental variability becomes inefficient. If the block size and shape is not appropriately chosen or if the block size is too large, the resulting experiment may not be a well controlled experiment in terms of variability and thus will provide inefficient results. Randomized block, Latin square and other complete block types of experiments are unsuitable for experiments in which large numbers of varieties or

treatments are used. They fail to adequately minimize the effect of soil heterogeneity (Lenter and Bishop 1993).

The randomised complete block design (RCBD), because of its simplicity continues to be a popular choice for many varietal trials. The precision of results relies heavily on the control of heterogeneity within blocks. Generally, the greater the heterogeneity within blocks, the poorer the precision of variety effect estimates. As the number of treatments increase, block size increases proportionally. This makes it difficult to maintain the homogeneity of experimental plots within the large blocks. The experimental error of a complete block design is generally expected to increase with the number of treatments. When the number of factors and or levels of factors increase, the number of treatment combinations increase very rapidly and it is not possible to accommodate all these treatment combinations in a single homogeneous block.

For a long time the methods used to overcome the difficulty of fitting a lot of treatments into one block of homogeneous units were; confounding one or more factorial contrasts with blocks or use split plot designs which in effect confound a factorial main effect. This reduction in size of block was achieved by sacrificing all or part of the information on certain treatment comparisons to achieve more precision on others (Idress and Khan, 2009).

In response to the need for efficient designs for a large number of treatments, Yates (1936) developed the group of incomplete block designs, known as quasi-factorials or lattices. As the name implies, each block in an incomplete block design, does not contain all treatments and a reasonably small block size can be maintained even if the number of treatments is large. With smaller blocks, the homogeneity of experimental units in the same block is easier

to maintain and a higher degree of precision can be generally achieved. Incomplete block designs or lattices divide each complete block into smaller blocks. These designs are arranged in blocks or groups that are smaller than a complete replication in order to eliminate heterogeneity (Yate, 1936)

Patterson and Williams (1976), extended Yate's method of construction to remove restrictions on the number of varieties and to generate generalized lattice designs, with widespread use, made of incomplete block design in variety trials. Generalised lattice designs are resolvable. If there is no gain in precision due to reduction in block size, these designs can be reanalyzed as if they were ordinary randomized complete blocks.

Bose and Nair (1939) presented a detailed account of construction of several incomplete block designs. Ma and Harrington (1948), during the period 1937 and 1946 used a total of 81 lattice designs of various kinds in Saskatoon, and Tisdale experiments at the University of Saskatchewan. The average increased efficiency of lattices over randomized blocks was 48%.

Lattice designs are now frequently used in the field of agriculture to test the yield of annual crops. A condition required in these designs is that the number of treatments used must be a perfect square such as 5^2 or 25, 6^2 or 36 etc. The most commonly used design is balanced lattice and partially balanced lattice design. The discrepancy between these two designs occurs on the number of replications to be used. Both require that the number of treatments must be a perfect square. In balanced lattice design, block size (k) is equal to the square root of the total number of treatments and the number of replications required is one more than the block size i.e. k+1. However in partially balanced lattice design any number of replications

can be used. If two replications are used in partially balanced lattice design, the design is called a simple lattice; with three replications, it is called a triple lattice; with four replications it is called a quadruple lattice, etc.

The advantages of lattice designs are:

- A large number of treatments may be compared within relatively small blocks (the incomplete blocks) and any number of treatments and replications can be used.
- Lattice designs provide a mechanism for better control of site variation and give a higher degree of precision.
- Lattice design may be analyzed as a randomized complete block design or completely randomized design, depending upon whether or not the incomplete blocks are arranged in complete blocks.

Patterson et al. (1978) suggested that if blocking is not done properly then lattice design can be analysed as RCBD by considering super blocks as ordinary blocks.

The disadvantages of lattice designs are:

- Analysis of lattice designs is more complex when missing plots occurs , covariance analysis is used or if the treatments are subjected to different error variances.
- Lattice designs are not available for all values of treatments, replications and incomplete block size.
- Lattice designs are more difficult to construct.

1.3 Statement of the Problem

Due to the decline of maize production in the country, its national consumption is over and above what is produced and therefore feeding the people will continue to be a major agricultural challenge. The country continues to rely on imports to meet deficits. Kenyan maize production averages 81kg per capita, significantly lower than the average demand of 103 kg per capita (Pingani, 2000). There is limited scope for expanding cultivated land under maize production since unutilized land is diminishing, degrading in soil fertility or unsuitable for maize production. Producing higher maize yield on existing cultivated land would be the best way of generating the extra maize grain to feed the nation. There is therefore a need to investigate the best maize variety which will give a higher yield given that the land holdings are constant while population growth is on the upward trend on yearly basis. Increasing Maize production in Kenya could be approached by planting the most appropriate maize varieties.

1.4 Broad objective

To determine the best performing maize variety in a maize variety trial conducted in two locations in Machakos County.

1.5 Specific objectives

i) To determine the best maize variety in each location

ii) To determine the overall best maize variety across the two locations

1.6 Hypothesis

Different maize varieties have the same mean yield

$$H_0 : \mu_1 = \mu_2 = \mu_3 ..... = \mu_k$$

against

H_1 at least one $\mu_i, _s$ is different

Where μ_k is the mean yield of the k^{th} variety.

In other words the null hypothesis is that all k varieties have the same mean yield.

The alternative hypothesis H_1 is that at least one of the treatments has a mean yield different from others.

1.7 Significance of the study

Effort to increase food production has been the key function in agricultural research institutions throughout the world. Such efforts have been spearheaded by Kenya Agricultural Research Institute (KARI). The Institute provides resources through field experimental trials which play momentous role in assembling, evaluating maize germ plasma and developing different varieties that are resistant to abiotic and biotic stresses. The study will provide information for immediate use to the agricultural extension field officers, small and large scale maize farmers. Secondly the study hope to provide useful findings to agricultural research bodies on the best varieties of maize that can be grown in arid and semi arid land (ASAL) for optimum yield.

1.8 Methodology

1.8.1 Study area

The trial was carried out in two locations; Katumani ($1^0$35'S; $37^0$14'E) and Kangundo ($1^0$18'S; $37^0$21' E). Kangundo is 57km on Eastern side of Nairobi while Katumani is about 80km south East of Nairobi and 8Km South of Machakos town along the Machakos-Wote road. Both locations are in Machakos county, Eastern Province at altitude from 1000 to 1600m above sea level with a semi- arid tropical climate described as agrecological zone (AEZ) IV with bimodal pattern of rainfall (Unesco 1974).

1.8.2 Experimental design.

The trial has 25 treatments (varieties). The experimental design used was a 5×5 partially balanced lattice design with 2 replications of 25 varieties each in 5 blocks. Each block contains 5 varieties. Twenty five varieties were randomly assigned to the experimental units in a randomized incomplete block design. The trial was carried out during the long rains season.

1.8.3 Field layout

All twenty five treatments (varieties) for both locations are shown in the layout in table 1, 2, 3 and 4.

Table 1: A 5×5 balanced-Lattice in the first Replicate for Katumani

Block 1	Block 2	Block 3	Block 4	Block 5
T9	T21	T2	T3	T22
T13	T8	T24	T10	T12
T18	T14	T5	T25	T20
T4	T11	T23	T6	T17
T1	T16	T7	T15	T19

Table 2: A 5×5 balanced-Lattice in the second Replicate for Katumani

Block 1	Block 2	Block 3	Block 4	Block 5
T9	T13	T18	T4	T1
T21	T8	T14	T11	T16
T2	T24	T5	T23	T7
T3	T10	T25	T6	T15
T22	T12	T20	T17	T19

Table 3 : A 5×5 balanced-Lattice in the first Replicate for Kangundo

Block 1	Block 2	Block 3	Block 4	Block 5
T9	T21	T2	T3	T22
T13	T8	T24	T10	T12
T18	T14	T5	T25	T20
T4	T11	T23	T6	T17
T1	T16	T7	T15	T19

Table 4 : A 5×5 balanced-Lattice in the second Replicate for Kangundo

Block 1	Block 2	Block 3	Block 4	Block 5
T9	T13	T18	T4	T1
T21	T8	T14	T11	T16
T2	T24	T5	T23	T7
T3	T10	T25	T6	T15
T22	T12	T20	T17	T19

1.8.4 Data collection

The data on the dried maize yield were recorded on data recording sheets. This raw data was keyed in the MS Excel spreadsheet and verified against the original data sheets. The maize yield data was expressed in tons per hectare.

1.8.5 Data Analysis

Data analysis was done using Microsoft Office Excel 2007, SAS (ver.9.1.3), Math type equations and R (ver.2.13.2) computer packages.

1.8.6 Assumption of analysis of variance

The following assumptions were made before carrying out an analysis of variance:

i) Each of the "n" (n=2) Populations i e. Katumani and kangundo are normally distributed with means, $\mu_1 = \mu_2$ and variances $\sigma_1^2 = \sigma_2^2$

ii) The two population variances are equal, $\sigma_1^2 = \sigma_2^2$ (i.e. there is homogeneity of variances).

iii) The independent samples are taken from 2 populations (equal and unequal).

iv) According to Annette (1990), a random variable X has a normal distribution with mean μ , variance σ and a probability density function of

$$f(x;\mu;\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma^2}\right)^2} \qquad \textbf{(1.1)}$$

This is denoted by $X \sim N(\mu, \sigma^2)$

$X \sim N(0,1)$ is called the standard normal distribution with $\mu = 0$ and $\sigma^2 = 1$.

CHAPTER 2

EXPLORATORY DATA ANALYSIS (EDA)

2.1 Introduction

Exploratory data analysis (EDA) is an approach used in analysing data sets and summarizing their main characteristics, often using visual methods. Exploratory data analysis (EDA) checks whether the data conforms to the underlying assumptions of a linear model before fitting a linear model. It gives multiple views of the data that may provide useful insights. EDA is a critical first step in analyzing the data from an experiment.

In this data set, EDA employed a variety of graphical techniques for the maize yield in Katumani and Kangundo areas. EDA employs graphical techniques and a few quantitative techniques. EDA employs a variety of techniques to: extract important variables; uncover underlying structure; maximize insight into a data set; detect outliers and anomalies; develop models and test underlying assumptions. These techniques are:

a) **Quantile - Quantile Plot (Q-Q plot)** is a plot of quantiles of quantitative response variables distribution against the quantiles of the normal distribution. If distribution is normal, the plot would have, for example, yield variable distributed closely around the straight line. Q-Q plots are generally used to determine whether the distribution of a variable matches the normal distribution. They allow detection of non-normality.

b) **Histogram** shows distribution and check normality. Yield data is normally distributed if it has a bell shape. The shape of the histogram indicates the closeness of the data to being normally distributed.

c) **Box plots** display two common measures of the variability or spread in the data set. They also show the outliers and extreme score either on upper or lower whisker.

2.2 EDA on Katumani

Katumani _replication 1

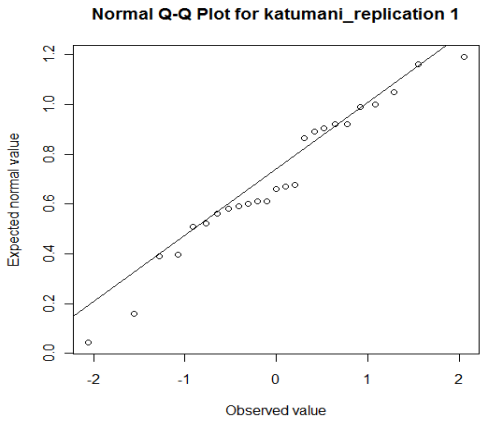

Figure 1

Figure 1 shows a Q-Q plot for Katumani maize yield data in replication 1. Most of the points are not in straight line or close to it. It shows deviation from normality on both ends as the points are far from the line. The plot indicates the existence of two clusters of the data. The yield variable does not match the test distribution (i.e. normal distribution).

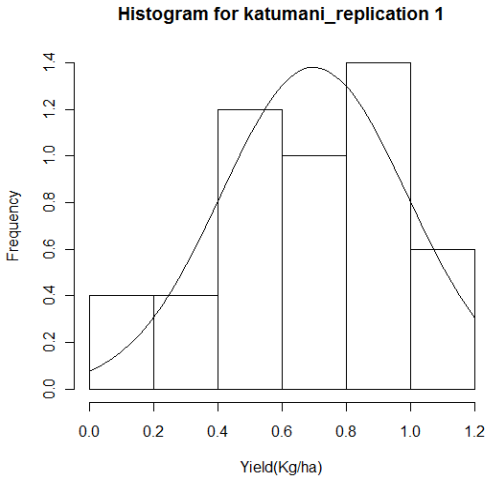

Figure 2

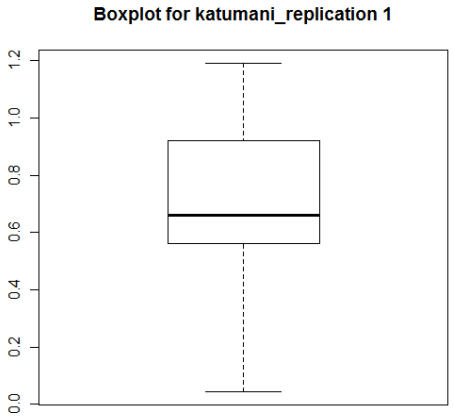

Figure 3

Figure 2 shows a histogram for Katumani maize yield data in replication 1 which indicates that the yield data is not normally distributed since it has no bell-shape. The distribution is

slightly left skewed. This is confirmed by a normal curve superimposed onto the histogram. The normal curve has a mean of 0.70 and standard deviation of 0.29.

Figure 3 shows a box plot for Katumani data in replication 1. The lower whisker is longer implying that the distribution is negatively skewed (or skewed towards left). The distribution is not normal as the median line does not divide the box equally .There are no outliers.

Katumani _replication 2

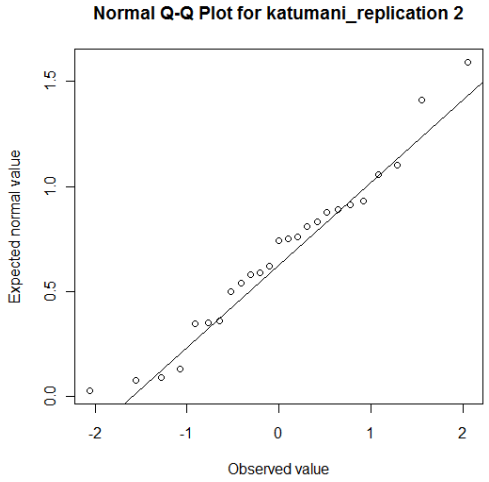

Figure 4

Figure 4 shows a Q-Q plot for Katumani maize yield data in replication 2.The distribution of the yield variable does not match the normal distribution since the points do not cluster around a straight line or close to it. The Q-Q plot shows deviation from normality on both sides i.e. upper and lower sides.

Figure 5 shows a histogram for Katumani maize yield data in replication 2 which indicates that the data is not normally distributed since it is not bell-shaped. This is confirmed by the

normal curve superimposed on the histogram which is positively skewed. The normal curve has a mean of 0.67 and standard deviation of 0.29.

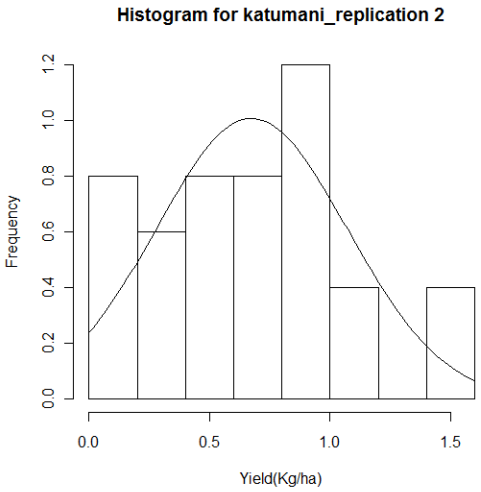

Figure 5

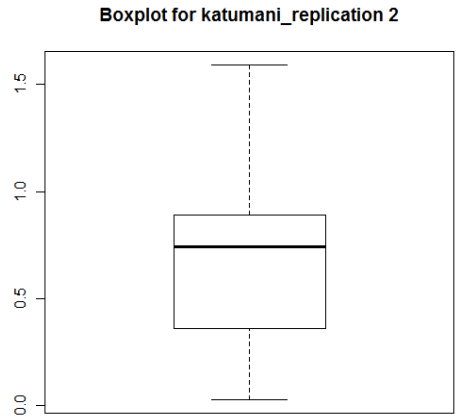

Figure 6

Figure 6 shows a box plot for Katumani maize yield data in replication 2. The upper whisker is longer implying that the underlying distribution is positively skewed. There are no outliers. The median line does not divide the box equally implying the data is not normally distributed.

Katumani _combined maize yield data

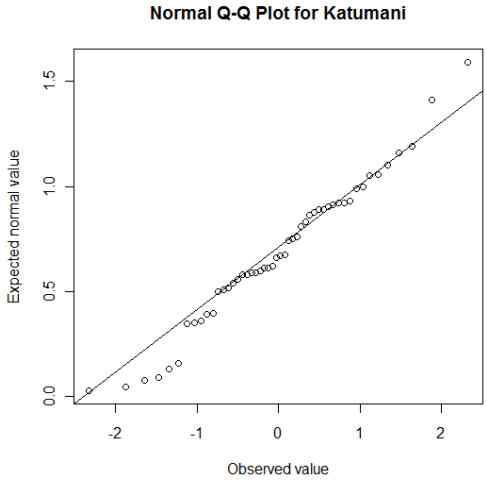

Figure 7

Figure 7 shows a Q-Q plot for Katumani combined maize yield data. Most of the data are closely around the straight line. There are outliers at lower and upper end. The plot slightly matches the test of normality.

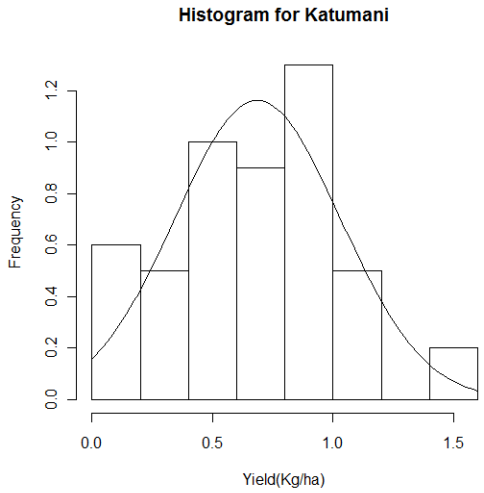

Figure 8

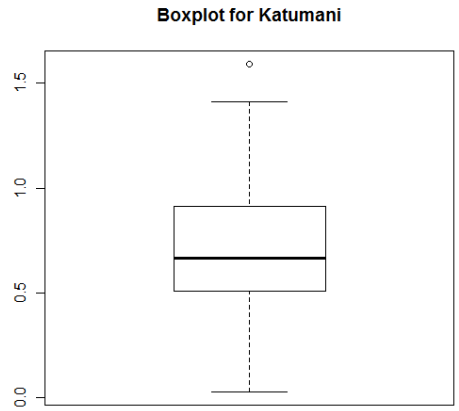

Figure 9

Figure 8 shows a histogram plot for Katumani combined maize yield data which have a bell shape implying that the yield data is symmetric and approximately normal. A normal curve superimposed onto the histogram confirms the variable is normally distributed. It has a mean yield of 0.687 and standard deviation of 0.343

Figure 9 shows a box plot for Katumani combined maize yield data. It shows one outlier on the upper whisker. The distribution is approximately normal as the median line is not far from the middle of the box and whiskers are almost of the same length.

2.3 EDA on Kangundo

Kangundo _replication 1

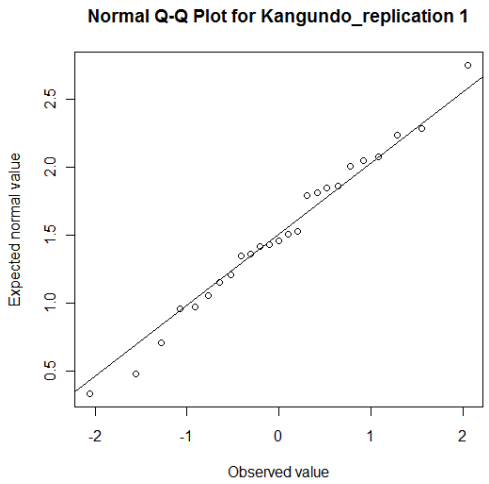

Figure 10

Figure 10 shows a Q-Q plot for Kangundo maize yield data in replication 1. The plot is linear thus ascertaining the underlying yield data distribution to be approximately normal because most of the points except a few cluster around the straight line.

Figure 11 in the next page shows a histogram for Kangundo maize yield data in replication 1 which is bell-shaped hence its underlying distribution is symmetrical and approximately normal. Normality is suggested by a normal curve superimposed on the histogram. The curve has a mean of 1.51 and standard deviation of 0.58

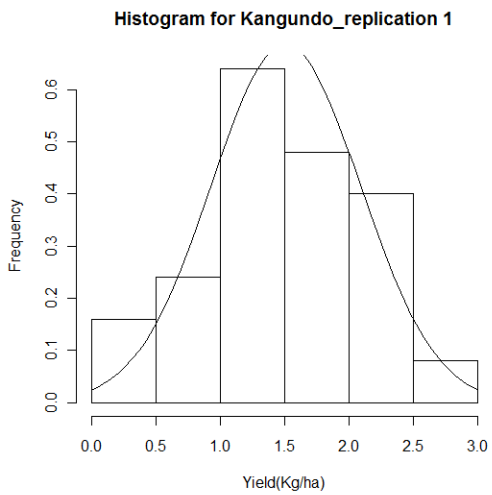

Figure 11

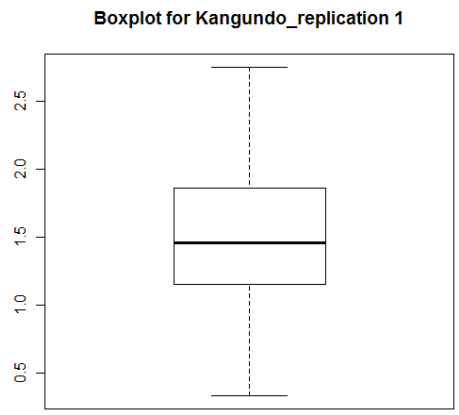

Figure 12

Figure 12 shows a box plot for Kangundo maize yield data in replication 1. The median line almost divides the box-plot equally hence the distribution is approximately normal. There are no outliers.

Kangundo _replication 1

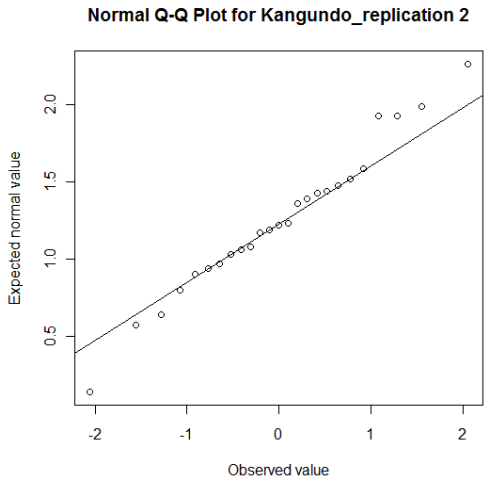

Figure 13

Figure 13 shows a Q-Q plot for Kangundo maize yield data in replication 2. The normal Q-Q plot is almost linear with few points deviating at the upper part thus the underlying distribution is approximately normal

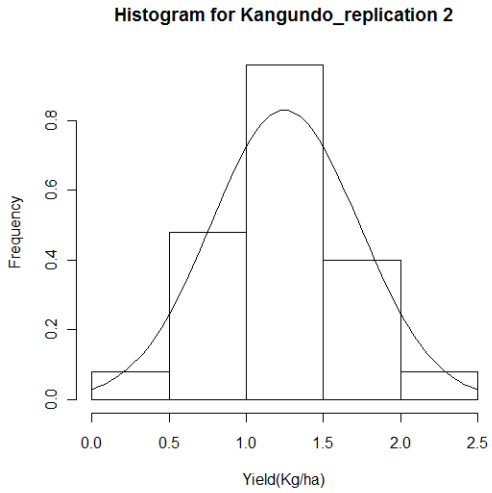

Figure 14

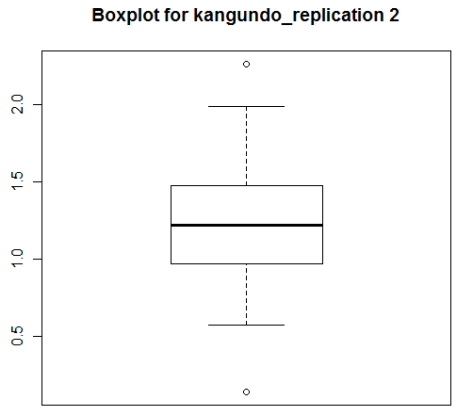

Figure 15

Figure 14 shows a histogram plot for Kangundo maize yield data in replication 2 which indicate that the yield data is normally distributed since it is bell-shaped. A normal curve superimposed onto the histogram reveals that the variable is normally distributed. The curve has a mean of 1.25 and standard deviation of 0.48

Figure 15 shows a box plot for Kangundo maize yield data in replication 2. The middle line is displayed at the middle of the box which confirms that the distribution is symmetric and normal. However there is an outlier on each side of the whiskers.

Kangundo _combined maize yield data

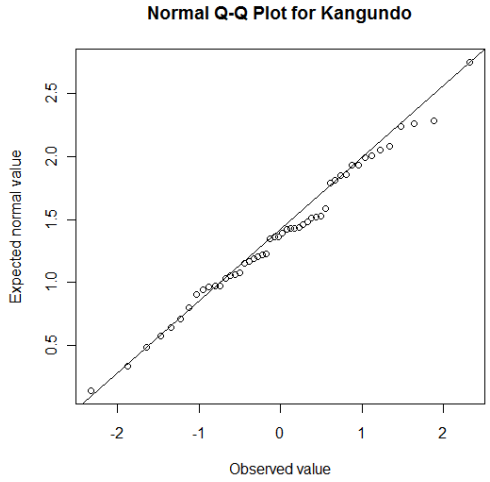

Figure 16

Figure 16 shows a Q-Q plot for Kangundo combined maize yield data which indicate that the data is normally distributed since they are closely around the straight line. There are outliers at the top and upper end. The plot matches the test of normality.

Figure 17 in the next page shows a histogram for Kangundo combined maize yield data which is bell-shaped implying the yield data distribution is symmetric and approximately normal. A normal curve superimposed onto the histogram suggests the variable is normally distributed. The curve has a mean yield of 1.378 and standard deviation of 0.543

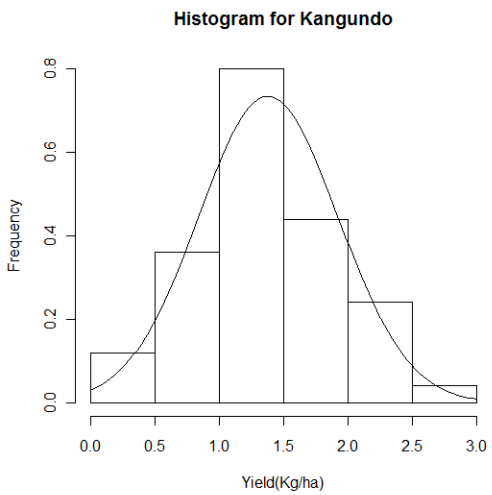

Figure 17

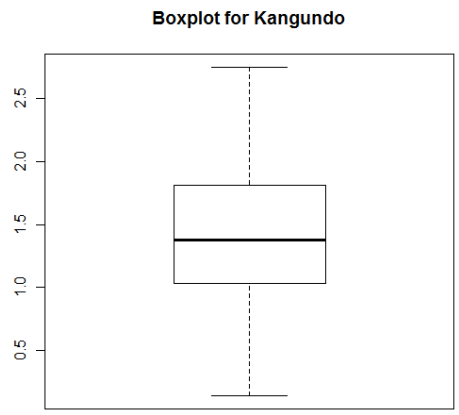

Figure 18

Figure 18 shows a box plot for Kangundo maize yield combined data. The distribution is almost symmetric and normal as the median line is not displayed far from the middle of the box.

2.4 Conclusion

From Exploratory Data Analysis (EDA), we can conclude that combined maize yield data over replications for both Katumani and Kangundo is normally distributed. The combined data for each location therefore confirms that the data conforms to the underlying assumptions of a linear model.

CHAPTER 3

ANALYSIS OF VARIANCE FOR SEPARATE LOCATIONS

3.1 Introduction

Analysis of Variance (ANOVA) is a statistical technique of using population means and variances to test uniformity or homogeneity of data. A population that is not homogeneous will have large variance while a homogeneous population will have small variance. The essence of ANOVA is that the total amount of variation in a set of data is split into two components; the amount which can be attributed to chance and the amount which can be attributed to specified causes. The basic principle of ANOVA is to test for the differences among the means of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between samples. This is done under the assumptions that the sampled populations are normally distributed.

3.1.1 Assumptions of ANOVA

i) Experimental errors are independently and normally distributed with mean zero and common variance.

ii) The effect of the i^{th} treatment remains same irrespective of the plot.

iii) The observations are independent.

iv) Parent population from which observations are taken is normal.

v) Various treatment and environmental effects are additive in nature.

3.2 Lattice Designs - Incomplete Blocks

Experimental designs are basically divided into two categories: complete block design and incomplete block design. Complete block design includes; completely randomized design (CRD), randomised complete block design (RCBD), latin square design and factorial design. Among these designs, RCBD is one of the most extensively used designs in agriculture. Incomplete block designs are designs in which only a subset of treatments are applied in each block. There are two types of incomplete block design; balanced incomplete block design (BIBD) and partially balanced incomplete block design (PBIBD).

In RCBD, a block should be homogeneous and each block contains a complete set of treatments. Therefore, a special requirement of RCBD experiments is that every block should contain observations on every treatment. However, when the number of treatments (t) is higher than the block size (k), every block cannot contain observations on every treatment.
In this case, an incomplete block design called quasi-factorials or lattice design is used instead of RCBD. Lattice designs, originally described by Yates (1936), are a special class of incomplete block designs used when number of treatments is large. A lattice design in a field trial involves grouping a block into smaller sub blocks. Each sub block cannot accommodate all the treatments. Grouping is done in such a way that every pair of treatments occurs together in the same block exactly once that is each pair of treatments occurs together in a block only once,($\lambda = 1$) .

In lattice design, the number of treatments (t) is a perfect square or a square of an integer, k such that $t = k^2$ .The design may be constructed for a number of treatments such as 9, 16, 25 etc. Lattice experiments require grouping blocks into replicates, with each replicate containing one observation for every treatment. This forces the number of blocks in each

replicate to be equal to the number of observations per block. That is, the number of blocks per replicate and the number of observations per block are both equal to $k = \sqrt{t}$. In addition, if the number of replicate groups (r) in Lattice experiments is equal to k + 1 then the design is referred to as a balanced lattice. If r is less than k + 1 then the design is referred to as a partially balanced lattice.

In constructing lattice designs of the balanced type, two fundamental relations are involved;

$$tr = kb \quad \textbf{(3.1)}$$

$$\lambda(t-1) = r(k-1) \quad \textbf{(3.2)}$$

Where $b \geq t$ and λ = the number of times (an integer) a treatment occurs with each of the other treatments within an incomplete block. If λ is equal for all pairs of treatments, the design is balanced. Lattice designs may also be used when the number of treatments is not a perfect square. Such cases are referred to as rectangular lattice designs. Lattice designs are very useful when comparing a large number of varieties because they correct heterogeneity.

3.3 Statistical model and ANOVA's format of a lattice design

The statistical model of a lattice design is given by:

$$Y_{ijkl} = \mu + R_i + B(R)_{j(i)} + T_k + e_{ijkl}$$

Where Y_{ijkl} = the observed value

μ = Overall mean yield.

R_i = Effect of the i^{th} replication

$B(R)_{j(i)}$ = Effect of the j^{th} block within the i^{th} replication.

T_k = Effect of the k^{th} treatment

e_{ijkl} = Random error

There are two ways of representing ANOVA table of lattice experiments. The different formats are shown in tables 5 and 6.

Table 5: The ANOVA table format for a lattice experiment at one location

Source of variation	Degree of freedom (df)	Sum of squares (SS)	Mean square $\left(MS = \frac{SS}{df}\right)$	Computed F
Replication (R)	$r-1$	SSR	MSR	$\frac{MSR}{MSE}$
Treatment (unadj.) (T)	k^2-1	SST (unadj.)	MST (unadj.)	$\frac{MST(unadj.)}{MSE}$
Blocks within replication (adj) (B)	$r(k-1)$	SSB (adj.)	MSB (adj.)	$\frac{MSB(unadj.)}{MSE}$
Intra-block error	$(k-1)(rk-k-1)$	SSE	MSE	
Total	rk^2-1	SST_0		

Table 6 : The ANOVA table format for a lattice experiment at one location with Adjustments

Source of variation	Degree of freedom (df)	Sum of squares (SS)	Mean square $\left(MS = \frac{SS}{df}\right)$	Computed F
Replication	$r-1$	SSR	MSR	$\frac{MSR}{MSE}$
Treatment (unadj.)	k^2-1	SST (unadj.)	MST (unadj.)	$\frac{MST(unadj.)}{MSE}$
Treatment (adj.)	k^2-1	SST (adj.)	MST (adj.)	$\frac{MST(adj.)}{MSE}$
Blocks within replication (adj)	$r(k-1)$	SSB (adj.)	MSB (adj.)	$\frac{MSB(adj.)}{MSE}$
Blocks within replication (unadj)	$r(k-1)$	SSB (unadj.)	MSB (unadj.)	$\frac{MSB(unadj.)}{MSE}$
Intra-block error	$(k-1)(rk-k-1)$	SSE	MSE	
Total	rk^2-1	SST_0		

The SAS ANOVA table format shown in table 7 below reflects two extra sources of variation which are not included in the standard format shown in tables 5 and 6. These are component B and randomized complete block error. Component B has similar features as blocks within replication while randomized complete block error is the sum of the blocks within replications sum of squares and the intra block error sum of squares. It is the appropriate error used if the experimental design uses a randomized complete block design (RCBD), with the replications taking the roles of complete blocks.

Table 7: The ANOVA table format for a Lattice Experiment at one Location given by SAS

Source	Degree of freedom (df)	Sum of squares (SS)	Mean square $\left(MS = \frac{SS}{df}\right)$
Replications	$r-1$	*SSR*	*MSR*
Blocks within Replication (adj)	$r(k-1)$	*SSB*	*MSB*
Component B	$r(k-1)$	*SSB*	*MSB*
Treatment (unadj.)	k^2-1	*SST*	*MST*
Intra-block error	$(k-1)(rk-k-1)$	*SSE*	*MSE*
Randomized Complete Block Error	$(r-1)(k^2-1)$	SSE_{RC}	MSE_{RC}
Total	rk^2-1	SST_0	

3.4 ANOVA of yield data for Katumani without adjustments

The total sum of squares (SST_0), replication sum of squares (SSR) and unadjusted treatment sum of squares (SSTunadj.) are obtained through SAS analysis and given in table 8

Table 8: Analysis of Variance for maize yield for Katumani Location

The SAS System
The Lattice Procedure
Analysis of Variance for yield

Source	df	Sum of Squares(SS)	Mean square(MS)
Replications	1	0.01656	0.01656
Blocks within Replication (adj)	8	0.3675	0.04593
Component B	8	0.3675	0.04593
Treatment (unadj.)	24	4.2517	0.1772
Intra-block error	16	0.8402	0.05251
Randomized Complete Block Error	24	1.2077	0.05032
Total	49	5.4760	0.1118

Additional Statistics for yield

Variance of Difference	0.0525
LSD at .01 Level	0.6409
LSD at .05 Level	0.4730
Efficiency Relative to RCBD	95.8241

Treatment Means
for yield

Treatment	Mean
1	0.3700
2	0.8350
3	0.6400
4	0.3700
5	0.8900
6	0.3200
7	0.3000
8	0.7250
9	1.1650
10	0.9000
11	0.0950
12	0.5700
13	0.5950
14	0.8100
15	1.0750
16	0.8225
17	0.5750
18	0.9100
19	0.0675
20	0.6600
21	0.7725
22	1.1000
23	1.0250
24	0.7100
25	0.7525

Treatment mean square of 0.1772 given from Anova table 8 will have to be adjusted due to block effects. The treatment means are not free from block effects as the numbers of treatments are high and therefore Anova will not provide a valid F test.

3.4.1 Adjustment of treatment means using adjustment factor, (a)

Treatment means are adjusted either up or down to remove any variation due to the block in which they occurred. In this way, all of the treatment means in the trial are compared on the same basis, without any bias due to local environmental variation in the field. Adjustment of treatment means also account for block to block variation within replications, so that treatments in different blocks are compared with precision.

Adjusted treatment means are used if:

a) The lattice design has a relative efficiency (RE) is greater than 100 percent compared to the RCBD i.e. $RE > 100\%$

b) Error due to blocks known as inter block error is greater than intra-block error i.e. blocks within replication mean square is greater than intra block error mean square ($E_b > E_e$)

c) A large number of treatments have been used and a significant difference among treatments may be expected.

Treatment means are adjusted by adjustment factor (a) is given by:

$$a = \frac{E_b - E_e}{k(r-1)E_b} \qquad \textbf{(3.3)}$$

Where E_b = adjusted inter block mean square ($MSB(adj.)$)

E_e = Intra-block mean square (MSE)

From the analysis we obtain:

i) E_b =0.04593 and E_e =0.05251. $E_b < E_e$ and therefore treatments in different blocks can be compared with equal precision as blocking has no effect.

ii) The relative efficiency is 95.82 % which is less than 100% and therefore adjustment is

necessary. However the efficiency obtained is not a mean achievement for a lattice design given the variations encountered in the ASALs.

A large number of treatments have been used and therefore the treatment means are to be adjusted for block effects. Adjusted sum of squares for treatments and unadjusted sum of squares for block are computed.

3.4.2 Computation of unadjusted block sum of squares (*SSB*(*unadj*.))

Unadjusted block sums of squares within replications i.e. SSB_1 and SSB_2 are computed first. The block totals B_1, B_2 …B_{10} for both replications are calculated and shown in tables 9 and 10.

Table 9: Arrangement of blocks and the treatments for Katumani location within the blocks and their totals

Replication 1

						Total
Block 1	9	13	18	4	1	
	1.19	0.6	0.89	0.395	0.39	3.465
Block 2	21	8	14	11	16	
	0.67	0.95	1	0.16	0.59	3.37
Block 3	2	24	5	23	7	
	0.92	0.66	0.865	1.16	0.52	4.125
Block 4	3	10	25	6	15	
	0.92	0.99	0.675	0.51	1.05	4.145
Block 5	22	12	20	17	19	
	0.61	0.56	0.58	0.61	0.045	2.405
Total	4.31	3.76	4.01	2.835	2.595	17.51

Table 10: Arrangement of blocks and the treatments for Katumani location within the blocks and their totals

	Replication 2					
						Total
Block 1	9	21	2	3	22	
	1.14	0.875	0.75	0.36	1.59	4.715
Block 2	13	8	24	10	12	
	0.59	0.5	0.76	0.81	0.58	3.24
Block 3	18	14	5	25	20	
	0.93	0.62	0.915	0.83	0.74	4.035
Block 4	4	11	23	6	17	
	0.345	0.03	0.89	0.13	0.54	1.935
Block 5	1	16	7	15	19	
	0.35	1.055	0.08	1.1	0.09	2.675
Total	3.355	3.08	3.395	3.23	3.54	16.6

Unadjusted blocks sum of square for replication 1 is

$$SSB_1(unadj.) = \frac{B_1^2 + B_2^2 .....B_5^2}{k} - \frac{R_1^2}{k^2} \quad \textbf{(3.4)}$$

$$= \frac{3.465^2 + 3.37^2 + 4.125^2 + 4.145^2 + 2.405^2}{5} - \frac{(15.51)^2}{25}$$

$$= 0.404756$$

Unadjusted blocks sum of square for replication 2 is

$$SSB_1(unadj.) = \frac{4.715^2 + 3.24^2 + 4.035^2 + 1.935^2 + 2.675^2}{5} - \frac{(16.6)^2}{25}$$

$$= 0.95958$$

Pooled unadjusted block sum of squares,

$$SSB(unadj.) = SSB_1(unadj.) + SSB_2(unadj.) \quad \textbf{(3.5)}$$

$$= 0.404756 + 0.95958$$

$$= 1.364336$$

3.4.3 Computation of adjusted treatment sum of squares ($SST(adj.)$) for block effects

Correction quantity (Q) is used to calculate adjusted sum of square for treatments $(SST(adj.))$ is given by (3.5)

$$Q = k(r-1)a\left[\left(\frac{r}{(r-1)(1+k\mu)}\right)(SSB(unadj.) - SSB(adj.))\right] \quad \textbf{(3.6)}$$

Where a = adjustment factor

E_b = adjusted inter block mean square ($MSB(adj.)$)

$= 0.04593$

E_e = Intra-block mean square (MSE)

$= 0.05251$

$k = 5, r = 2$

Substituting E_b, E_e, k and r in (3.3)

$$a = \frac{0.04593 - 0.0525}{5(2-1)0.04593}$$

$= -0.027$

Substituting $k, r, a, SSB(unadj)$ and $SSB(adj)$ in (3.5)

$$Q = 5(1)(-0.027)\left[\left(\frac{2}{(2-1)(1+5(-0.027))}\right)(1.3643 - 0.3675)\right]$$

$= -0.311$

Quantity Q is subtracted from the unadjusted treatment sum of squares to obtain the adjusted sum of squares for treatment i.e.

$$SST(adj.) = SST(unadj.) - Q \quad \textbf{(3.7)}$$

$= 4.2517 - (-0.3111)$

$= 4.5628$

3.5 ANOVA of yield data for Katumani with adjustments

Adjusted treatment sum of squares, $(SST(adj.))$ unadjusted block sum of squares $(SSB(unadj.))$ values obtained in section 3.4.2 are entered in the Anova table 11.

Mean square of adjusted treatment, $MST(adj.) = \frac{4.5628}{24}$

$= 0.1901$

Mean square of unadjusted block, $MSB(unadj.) = \frac{1.3643}{8}$

$= 0.1705$

These values are entered in table 11

Table 11: Anova table for maize yield for Katumani Location with adjustments

Source of variation	Degree of freedom (df)	Sum of squares (SS)	Mean square $\left(MS = \frac{SS}{df}\right)$	F_{calc}	F_{tab}
Replication	1	0.01656	0.01656	0.32	4.49
Treatment (unadj.)	24	4.2517	0.1772	3.37*	2.24
Treatment (adj.)	24	4.5628	0.1901	3.62*	2.24
Blocks within replication (adj)	8	0.3675	0.04593	0.87	2.59
Blocks within replication (unadj)	8	1.3643	0.1705	3.25*	2.59
Intra-block error	16	0.8402	0.05251		
Total	49	5.4760	0.1118		

There is significant difference among unadjusted treatment means at 5% level of significance since F computed > F tabulated. After treatment means were adjusted, the situation still remained the same i.e. adjusted treatment means are significant at 5% level of significance.

Adjusted block within replications is not significant at 5% level of significance while unadjusted blocks within replications are significant at the same level.

3.5.1 Computation of adjusted block total C_b values

An adjusted block C_b value of a block is the difference between column total of replication 2 and their corresponding block total of replication 1.

Example:

C_1 value of the 1st block in table 12 = column total of replication 2 - block total of replication 1

= 3.355 - 3.465

= -0.11

The column total of replication 2 in table 12 is obtained from the total of each column of replication 2 in table 10. The block total of replication 1 in table 12 is obtained from the total of each block of replication 1 in table 9.

Table 12: Computation of C_b values for blocks in Katumani Replication 1

Block	Column total of Replication 2	Block total of Replication 1	C_b value	
1	3.355	3.465	-0.11	(C_1)
2	3.08	3.37	-0.29	(C_2)
3	3.395	4.125	-0.73	(C_3)
4	3.23	4.145	-0.915	(C_4)
5	3.54	2.405	1.135	(C_5)
Total	16.6	17.51	-0.91	(R_{C1})

Total C_b value (R_{C1}) of a replication is the sum of all individual C_b values in that replication.

Total C_b value (R_{C1}) for replication 1, $R_{C1} = \sum_{i=1}^{5} C_i$

$$= -0.11 + (-0.29) + (-0.73) + 1.135$$

$$= -0.91 \text{ (table 12)}$$

Total C_b value (R_{C2}) for replication 2, $R_{C2} = \sum_{i=1}^{5} C_i$

$$= 0.91 \text{ (table 13)}$$

Total C_b values for both replications are used to:

i) Check whether arithmetic calculations have been done correctly. If correct, summation of C_b values for both replications should be zero.

Example, for this case, $Total = R_{C1} + R_{C2}$

$$= -0.91 + 0.91 = 0$$

ii) Compute correction value, μC_b (section 3.5.2)

For katumani replication 2, C_b values shown in table 13 are given by subtracting block total in replication 1 from the corresponding column total in replication 2.

Table 13: Computation of C_b values for blocks in Katumani Replication 2

Block	Column total of Replication 2	Block total of Replication 1	C_b values	
1	4.31	4.715	-0.405	(C_6)
2	3.76	3.24	0.52	(C_7)
3	4.01	4.035	-0.025	(C_8)
4	2.835	1.935	0.9	(C_9)
5	2.595	2.675	-0.08	(C_{10})
Total	17.51	16.6	0.91	(R_{C2})

3.5.2 Computation of correction value, aC_b

A correction term for each block is computed by multiplying each C_b value by the quantity $a = -0.027$, given by (3.3)

For replication 1, these values are:

$$aC_1 = -027 \times -0.11 = 0.00297 = 0.003$$

$$aC_2 = 0.008,\ aC_3 = 0.020,\ aC_4 = 0.025,\ aC_5 = -0.031$$

For replication 2, these values are:

$$aC_6 = 0.011,\ aC_7 = -0.014,\ aC_8 = 0.001,\ aC_9 = -0.024,\ aC_{10} = 0.002$$

The total sum of all C_b values should add up to zero i.e.

$$\text{Total sum of } C_b \text{ values} = \sum_{i=1}^{10} aC_i = aC_1 + aC_2 + ... + aC_{10} = 0 \qquad \textbf{(3.8)}$$

$$= 0.03 + 0.008 + ... + 0.002 = 0$$

The aC_b Values of replication 1 are entered along the last column of table 14 as shown while aC_b values for replication 2 are entered along the last row in the same table.

Table 14: Treatment totals and correction values for Katumani

Block 1	9	13	18	4	1	aC_1=
	2.33	1.19	1.82	0.74	0.74	0.003
Block 2	21	8	14	11	16	aC_2=
	1.545	1.45	1.62	0.19	1.645	0.008
Block 3	2	24	5	23	7	aC_3=
	1.67	1.42	1.78	2.05	0.6	0.020
Block 4	3	10	25	6	15	aC_4=
	1.28	1.8	1.505	0.64	2.15	0.025
Block 5	22	12	20	17	19	aC_5=
	2.2	1.14	1.32	1.15	0.135	-0.031
	aC_6=	aC_7=	aC_8=	aC_9=	aC_{10}=	
	0.011	-0.014	0.001	-0.024	0.002	

3.5.3 Adjusted treatment totals and adjusted treatment means

Each treatment total in table 14 is adjusted for block effects by applying the block corrections appropriate to the blocks in which that treatment appears.

For example

The adjusted treatment total for treatment 9 in table $14 = T9 - aC_1 - aC_6$

$$= 2.33 - 0.003 - 0.011$$

$$= 2.316$$

Table 15 is constructed to show the treatment total adjusted for block effects.

Table 15: Adjusted treatment totals for katumani

Block 1	9	13	18	4	1
	2.316	1.201	1.816	0.761	0.735
Block 2	21	8	14	11	16
	1.526	1.456	1.611	0.206	1.635
Block 3	2	24	5	23	7
	1.639	1.414	1.760	2.055	0.578
Block 4	3	10	25	6	15
	1.244	1.789	1.480	0.640	2.123
Block 5	22	12	20	17	19
	2.220	1.185	1.350	1.205	0.163

Adjusted treatment means are shown in table 16 obtained by dividing each value in table 15 by two since each total contains two observations from 2 replications.

Table 16: Adjusted treatment means for katumani

Block 1	9	13	18	4	1
	1.158	0.601	0.908	0.381	0.367
Block 2	21	8	14	11	16
	0.763	0.728	0.806	0.103	0.818
Block 3	2	24	5	23	7
	0.820	0.707	0.880	1.027	0.289
Block 4	3	10	25	6	15
	0.622	0.895	0.740	0.320	1.062
Block 5	22	12	20	17	19
	1.110	0.592	0.675	0.602	0.082

3.5.4 Computation of effective mean square (effective error variance), $E_e^{'}$

There are differences between adjusted treatment totals (table 15) and treatment totals (table-14) and therefore it is necessary to compute effective mean square or effective error variance, $E_e^{'}$ which is given by:

$$E_e^{'} = \left(1 + \frac{rka}{k+1}\right) E_e \qquad \textbf{(3.9)}$$

$$E_e^{'} = \left(1 + \frac{(2)(5)(-0.027)}{5+1}\right) 0.05251$$

$$= 0.0501 \quad \text{where } a = -0.027 \text{ from (3.3)}$$

$E_e = 0.05251$ (table 8), $r = 2$ and $k = 5$

According to Cochran and Cox, 1950 effective error mean square ($E_e^{'}$) is used in the denominator of F ratio test instead of E_e to test if there are differences among adjusted treatment means.

F ratio test of adjusted means $F_{calc(0.05,24,16)} = \dfrac{MST(adj.)}{E_e^{'}}$

$$= \frac{0.1901}{0.0501}$$

$$= 3.79$$

$$F_{tab(0.05,24,16)} = 2.24$$

Since $F_{calc} > F_{tab}$, adjusted treatment means are significantly different at 5% level of significance.

3.6 Comparison of treatment means and the least significant difference (LSD) test

In a partially balanced lattice design, treatments that occur in the same block are compared with greater precision i.e. smaller standard error than the treatments that occur in different blocks. Standard error for comparing any two treatment means that occur together in the same block is given by; $SE(d_1) = \sqrt{\frac{2E_e}{r}\left[1+\left(r-1\right)a\right]}$ **(3.10)**

Standard error for comparing any two treatments in the same block in Katumani location is

$$SE(d_1) = \sqrt{\frac{2(0.053)}{2}\left[1+(-0.027)\right]}$$

$$= 0.227$$

where $a = -0.027$, $r = 2$ and $E_e = 0.053$

The formula for determining standard error for comparing treatment means that occur in different blocks is given by; $SE(d_2) = \sqrt{\frac{2E_e}{r}\left(1+ra\right)}$ **(3.11)**

Standard error for comparing any two treatments in different blocks in Katumani location is

$$SE(d_2) = \sqrt{\frac{2(0.053)}{2}\left[1+2(-0.027)\right]}$$

$$= 0.224$$

These standard errors when multiplied by the tabular *t* value for the intra-block error degrees of freedom at the specified level of significance will provide LSD value with which the adjusted treatment means can be compared for significant differences.

The LSD test is the simplest of the procedures for making pairwise comparisons. The procedure provides for a single LSD value, at a prescribed level of significance, which serves as the boundary between significant and non significant difference between any pair of

treatment means. That is, two treatments are declared significantly different at a prescribed level of significance α, if their difference exceeds the computed LSD value i.e.

$$\overline{x_1} - \overline{x_2} > LSD_\alpha \qquad \textbf{(3.12)}$$

Otherwise they are not considered significantly different.

The aim is to determine the best performing maize variety through comparison of two treatment means in two cases:

Case 1: When treatments are in the same block

For two treatments, take $T_9 = \overline{x_1} = 1.158$ and $T_{18} = \overline{x_2} = 0.908$

$$\overline{x_1} - \overline{x_2} = T_9 - T_{18}$$

$$= 1.158 - 0.908$$

$$= 0.250$$

Treatments are from the same block; $SE(d_1) = 0.227$ from (3.10)

Let the level of significance, $\alpha = 5\%$, $df = 16$

$$LSD_\alpha = \left(t_{df,\alpha/2}\right)\left(SE(d_1)\right) \qquad \textbf{(3.13)}$$

From t-table; t tabular = $t_{df,0.05} = t_{16,0.025} = 2.583$

$$LSD_\alpha = 2.583 \times 0.227$$

$$= 0.586$$

$\overline{x_1} - \overline{x_2} < LSD_{0.05}$ hence the two treatment means are not significantly different.

Case 2: When treatments are in different blocks

For two treatments, take $T_9 = \overline{x_1} = 1.158$ and $T_{22} = \overline{x_2} = 1.110$

$$\overline{x_1} - \overline{x_2} = T_9 - T_{22}$$

$$= 1.158 - 1.110$$

$$= 0.048$$

Treatments are from different blocks; $SE(d_2) = 0.224$ from (3.11)

Let the level of significance, $\alpha = 5\%$, $df = 16$

$$LSD_{\alpha} = \left(t_{df,\alpha/2}\right)\left(SE(d_2)\right) \quad \textbf{(3.14)}$$

From t-table; t tabular = $t_{df,0.05} = t_{16,0.025} = 2.583$

$$LSD_{\alpha} = 2.583 \times 0.224$$

$$= 0.579$$

$\overline{x_1} - \overline{x_2} < LSD_{0.05}$ hence two treatment means are not significantly different.

The best performing maize variety can be determined by comparing the highest yielding maize variety with the rest to find out whether there are some significant differences as shown in table 17.

Table 17: Comparison of treatment means with T9

Treat no.	$\overline{X_i}$	$\overline{X_i}$	$\overline{X_1} - \overline{X_i}$	LSD	Signif
T9	X1	1.158	0.000		
T22	X2	1.110	0.048	0.579	NS
T15	X3	1.062	0.096	0.579	NS
T23	X4	1.027	0.131	0.579	NS
T18	X5	0.908	0.250	0.586	NS
T10	X6	0.895	0.263	0.579	NS
T5	X7	0.880	0.278	0.579	NS
T2	X8	0.820	0.338	0.579	NS
T16	X9	0.818	0.341	0.579	NS
T14	X10	0.806	0.352	0.579	NS
T21	X11	0.763	0.395	0.579	NS
T25	X12	0.740	0.418	0.579	NS
T8	X13	0.728	0.430	0.579	NS
T24	X14	0.707	0.451	0.579	NS
T20	X15	0.675	0.483	0.579	NS
T3	X16	0.622	0.536	0.579	NS
T17	X17	0.602	0.556	0.579	NS
T13	X18	0.601	0.558	0.586	NS
T12	X19	0.592	0.566	0.579	NS
T4	X20	0.381	0.777	0.586	*
T1	X21	0.367	0.791	0.586	*
T6	X22	0.320	0.838	0.579	*
T7	X23	0.289	0.869	0.579	*
T11	X24	0.103	1.055	0.579	*
T19	X25	0.082	1.076	0.579	*

NS = Not significant

* = Significant at 5% level.

Treatments T4, T1, T6, T7, T11 and T19 are significantly different from T9 at 5% level of significance. Other treatment means are not significantly different from T9 at 5% level of significance and therefore different types of trials should be conducted to investigate other factors.

3.7 ANOVA of yield data for Kangundo

Through SAS analysis, Anova table for kangundo is obtained and shown as table 18.

Table 18: Analysis of Variance for maize yield for Kangundo Location

The SAS System
The Lattice Procedure

Analysis of Variance for yield

Source	df	Sum of Squares(SS)	Mean square(MS)
Replications	1	0.4869	0.4869
Blocks within Replication (adj)	8	2.2871	0.2859
Component B	8	2.2871	0.2859
Treatments (unadj.)	24	3.2770	0.1365
Intra-block error	16	6.8801	0.4300
Randomized Complete Block Error	24	9.1673	0.3820
Total	49	12.9312	0.2639

Additional Statistics for yield

Variance of Difference	0.4300
LSD at .01 Level	1.8341
LSD at .05 Level	1.3534
Efficiency Relative to RCBD	88.8284

Treatment Means
for yield

Treatment	Mean
1	1.7750
2	1.7050
3	1.4275
4	0.6750
5	1.7100
6	1.4450
7	1.2125

8	1.5750
9	1.4250
10	1.5450
11	1.4700
12	1.4200
13	0.8875
14	1.6400
15	1.2925
16	1.6550
17	1.1600
18	1.4100
19	1.3220
20	1.2800
21	1.0925
22	1.3725
23	1.6650
24	1.6200
25	1.4000

Treatment means are not free from block effects because the numbers of treatments used are many. The analysis of variance will not provide a valid F test and therefore adjustments are needed on treatment mean square.

Blocking has no effect as E_b = 0. 2859 is less than E_e =0.4300 i.e. $E_b < E_e$. Efficiency relative to RCBD is 88.8% and adjustments of treatments are necessary as the number of treatments used is large.

3.7.1 Adjustment of treatment means using adjustment factor, (a)

Unadjusted block sums of squares within replications i.e. SSB_1 and SSB_2 are computed first. The block totals B_1, B_2 …B_{10} for both replications are calculated and shown in next two tables 19 and 20.

Table 19: Arrangement of blocks and the treatments for Kangundo location within the blocks and their totals

Kangundo replication 1

						Total
Block 1	9	13	18	4	1	
	1.46	0.335	1.85	0.71	2.75	7.105
Block 2	21	8	14	11	16	
	1.155	1.79	1.35	1.51	2.08	7.885
Block 3	2	24	5	23	7	
	2.24	2.05	1.43	1.81	1.205	8.735
Block 4	3	10	25	6	15	
	2.285	2.01	1.86	0.96	1.525	8.64
Block 5	22	12	20	17	19	
	0.48	1.36	0.97	1.42	1.054	5.284
Total	7.62	7.545	7.46	6.41	8.614	37.649

Table 20: Arrangement of blocks and the treatments for Kangundo location within the blocks and their totals.

Kangundo replication 2

						Total
Block 1	9	21	2	3	22	
	1.39	1.03	1.17	0.57	2.265	6.425
Block 2	13	8	24	10	12	
	1.44	1.36	1.19	1.08	1.48	6.55
Block 3	18	14	5	25	20	
	0.97	1.93	1.99	0.94	1.59	7.42
Block 4	4	11	23	6	17	
	0.64	1.43	1.52	1.93	0.9	6.42
Block 5	1	16	7	15	19	
	0.8	1.23	1.22	1.06	1.59	5.9
Total	5.24	6.98	7.09	5.58	7.825	32.715

Unadjusted blocks sum of square for replication 1 is

$$SSB_1(unadj.) = \frac{B_1^{\,2} + B_2^{\,2} + ... + B_5^{\,2}}{k} - \frac{R_1^{\,2}}{k^2} \qquad \textbf{(3.4)}$$

$$= \frac{7.105^2 + 7.885^2 + 8.735^2 + 8.64^2 + 5.284^2}{5} - \frac{(37.649)^2}{25}$$

$$= 1.607$$

Unadjusted blocks sum of square for replication 2 is

$$SSB_1(unadj.) = \frac{6.425^2 + 6.55^2 + 7.42^2 + 6.42^2 + 5.9^2}{5} - \frac{(32.715)^2}{25}$$

$$= 0.242$$

Pooled unadjusted block sum of squares,

$$SSB(unadj.) = SSB_1(unadj.) + SSB_2(unadj.) \quad \textbf{(3.5)}$$

$$= 1.607 + 0.242$$

$$= 1.849$$

3.7.2 Computation of adjusted treatment sum of squares ($SST(adj.)$) for block effects

Correction quantity Q is used to calculate adjusted sum of squares for treatments i.e. $SST(adj.)$ is given by (3.6)

$$Q = k(r-1)a\left[\left(\frac{r}{(r-1)(1+k\mu)}\right)(SSB(unadj.) - SSB(adj.))\right] \quad \textbf{(3.6)}$$

E_b = adjusted inter block mean square ($MSB(adj.)$)

$= 0.2859$

E_e = Intra-block mean square (MSE)

$= 0.4300$

$k = 5, r = 2$

Substituting E_b, E_e, k and r in (3.3)

$$a = \frac{0.2859 - 0.4300}{5(2-1)0.2859}$$

$$= -0.101$$

Substituting $k, r, a, SSB(unadj)$ and $SSB(adj)$ in (3.6)

$$Q = 5(1)(-0.101)\left[\left(\frac{2}{(2-1)(1+5(-0.101))}\right)(1.849-2.2871)\right]$$

$$= -0.894$$

Quantity Q is subtracted from the unadjusted treatment sum of squares to obtain the adjusted sum of squares for treatment i.e.

$$SST(adj.) = SST(unadj.) - Q \qquad \textbf{(3.7)}$$

$$= 3.2770 - (-0.894)$$

$$= 2.383$$

3.8 ANOVA of yield data for kangundo with adjustments

Adjusted treatment sum of squares, $(SST(adj.))$ and unadjusted block sum of squares $(SSB(unadj.))$ values obtained in section 3.71 are entered in Anova table 21.

Mean square of adjusted treatment, $MST(adj.) = \frac{2.383}{24}$

$$= 0.0993$$

Mean square of unadjusted block, $MSB(unadj.) = \frac{1.849}{8}$

$$= 0.1705$$

Table 21: Anova table for maize yield for Kangundo Location with adjustments

Source of variation	Degree of freedom (df)	Sum of squares (SS)	Mean square $\left(MS = \frac{SS}{df}\right)$	F_{calc}	F_{tab}
Replication	1	0.4869	0.4869	1.13	4.49
Treatment (unadj.)	24	3.2770	0.1365	0.32	2.24
Treatment (adj.)	24	2.383	0.0993	0.23	2.24
Blocks within replication (adj)	8	2.2871	0.2859	0.66	2.59
Blocks within Replication (unadj)	8	1.849	0.2311	0.54	2.59
Intra-block error	16	6.8801	0.4300		
Total	49	5.4760	0.1118		

Unadjusted treatment means are not significantly different at 5% level of significance since F computed < F tabulated. After treatment means were adjusted, the situation still remained the same i.e. adjusted treatment means were not significantly different at 5% level of significance. Both adjusted and unadjusted block within replications are not significantly different at 5 % level of significance.

3.8.1 Computation of adjusted block total C_b values

For Kangundo replication 1, C_b values shown in table 22 are given by subtracting block total in replication 1 from the corresponding column total in replication 2.

For example, C_1 value of the 1^{st} block $= 5.24 - 7.105$

$$= -1.865$$

Table 22: Computation of C_b values for blocks in Kangundo Replication 1

Block	Column total of Replication 2	Block total of Replication 1	C_b value	
1	5.24	7.105	-1.865	(C_1)
2	6.98	7.885	-0.905	(C_2)
3	7.09	8.735	-1.645	(C_3)
4	5.58	8.64	-3.06	(C_4)
5	7.825	5.284	2.541	(C_5)
Total	32.715	37.649	-4.934	(R_{C1})

For Kangundo replication 2, C_b values shown in table 23 are obtained by subtracting block total in replication 1 from the corresponding column total in replication 2.

For example, for block 1; C_1 value of the 1st block $= 7.62 - 6.425$

$= 1.195$

Table 23: Computation of C_b values for blocks in Kangundo Replication 2

Block	Column total of Replication 2	Block total of Replication 1	C_b value	
1	7.62	6.425	1.195	(C_1)
2	7.545	6.55	0.995	(C_2)
3	7.46	7.42	0.04	(C_3)
4	6.41	6.42	-0.01	(C_4)
5	8.614	5.9	2.714	(C_5)
Total	37.649	32.715	4.934	(R_{C2})

Total value of C_b for replication 1 and 2 are obtained and add up to zero indicating that arithmetic calculation has been done correctly.

Total of C_b values for replication I is $= R_{C1} = -4.934$

Total of C_b values for replication 2 is $= R_{C2} = 4.934$

$$Total = R_{C1} + R_{C2} = 0$$

3.8.2 Computation correction value, aC_b

A correction term for each block is computed by multiplying each C_b value by the quantity $a = -0.101$ given by (3.3) in sub section 3.7.1

For replication 1, these values are:

$aC_1 = -0.101 \times -1.865 = 0.188$, since $C_1 = -1.865$ from table 22

$aC_2 = 0.091, aC_3 = 0.166, aC_4 = 0.309, aC_1 = -0.257$

For replication 2, these values are:

$aC_6 = -0.121, aC_7 = -0.100, aC_8 = 0.004, aC_9 = -0.001, aC_{10} = -0.274$

Total C_b values = $\sum_{i=1}^{10} aC_i = aC_1 + aC_2 + ... + aC_{10} = 0$ **(3.15)**

$$= 0.188 + 0.091 + ... + (-0.274) = 0$$

The aC_b Values of replication 1 are entered along the last column of table 24 as shown while aC_b values for replication 2 are entered along the last row in the same table.

Table 24: Treatments totals and correction values for Kangundo

Block 1	9 2.85	13 1.775	18 2.82	4 1.35	1 3.55	aC_1= 0.188
Block 2	21 2.185	8 3.15	14 3.28	11 2.94	16 3.31	aC_2= 0.091
Block 3	2 3.41	24 3.24	5 3.42	23 3.33	7 2.425	aC_3= 0.166
Block 4	3 2.855	10 3.09	25 2.8	6 2.89	15 2.585	aC_4= 0.309
Block 5	22 2.745	12 2.84	20 2.56	17 2.32	19 2.644	aC_5= -0.257
	aC_6= -0.121	aC_7= -0.100	aC_8= -0.004	aC_9= 0.001	aC_{10}= -0.274	

38.3 Adjusted treatment totals and adjusted treatment means

Each treatment in table 24 is adjusted for block effects by applying the block corrections appropriate to the blocks in which that treatment appears.

For example

The adjusted treatment total for T9 in table 22 $= T9 - aC_1 - aC_6$

$= 2.85 - 0.188 - (-0.121)$

$= 2.783$

Table 25 is constructed showing the treatment total adjusted for block effects.

Table 25: Adjusted treatment totals for kangundo

Block 1	9	13	18	4	1
	2.783	1.687	2.636	1.161	0.718
Block 2	21	8	14	11	16
	2.215	3.159	3.193	2.848	3.493
Block 3	2	24	5	23	7
	3.365	3.174	3.258	3.163	2.533
Block 4	3	10	25	6	15
	2.667	2.881	2.495	2.580	2.550
Block 5	22	12	20	17	19
	3.123	3.197	2.821	2.576	3.175

Adjusted treatment means are obtained by dividing each value in table 25 by two since each total contains two observations from 2 replications.

Table 26: Adjusted treatment means for kangundo

Block 1	9	13	18	4	1
	1.391	0.844	1.318	0.580	0.359
Block 2	21	8	14	11	16
	1.107	1.580	1.597	1.424	1.747
Block 3	2	24	5	23	7
	1.682	1.587	1.629	1.581	1.267
Block 4	3	10	25	6	15
	1.333	1.441	1.248	1.290	1.275
Block 5	22	12	20	17	19
	1.561	1.599	1.411	1.288	1.588

3.8.4 Computation of effective mean square (effective error variance), $E_e^{'}$

There are differences between adjusted treatment totals (table 24) and unadjusted treatment totals (table 25) and therefore it is necessary to compute effective mean square or effective error variance, $E_e^{'}$ which is given by:

$$E_e^{'} = \left(1 + \frac{rka}{k+1}\right) E_e \qquad \textbf{(3.16)}$$

$$E_e^{'} = \left(1 + \frac{(2)(5)(-0.101)}{5+1}\right) 0.4300$$

$$= 0.3576 \quad \text{where } a = -0.101 \text{ from (3.3)}$$

$E_e = -0.4300$ (table 15), $r = 2$ and $k = 5$

F ratio test of adjustment means $F_{calc(0.05,24,16)} = \dfrac{MST(adj.)}{E_e^{'}}$

$$= \frac{0.0993}{0.3576}$$

$$= 0.2777$$

$$F_{tab(0.05,24,16)} = 2.24$$

Since $F_{calc} < F_{tab}$, adjusted treatment means are not significantly different at 5% level of significance.

3.9 Comparison of treatment means and LSD

Standard error for comparing any two treatment means that occur together in the same block is given by; $SE(d_1) = \sqrt{\dfrac{2E_e}{r}\left[1 + (r-1)a\right]}$ **(3.17)**

Standard error for comparing any two treatments in the same block in Kangundo location is

$$SE(d_1) = \sqrt{\frac{2(0.43)}{2}(1 - 0.101)}$$

$= 0.623$

where $a = -0.101, r = 2$ and $E_e = 0.4300$

The standard error for comparing treatment means that occur in different blocks is given by;

$$SE(d_2) = \sqrt{\frac{2E_e}{r}(1+ra)} \quad \textbf{(3.18)}$$

Standard error for comparing any two treatments in different blocks in Kangundo location is

$$SE(d_2) = \sqrt{\frac{2(0.430)}{2}[1+2(-0.101)]}$$

$= 0.586$

Comparison test is done to determine the best maize variety in two cases:

Case 1: When treatments are in the same block

For two treatments, take $T_{16} = \overline{x_1} = 1.747$ and $T_{22} = \overline{x_5} = 1.597$

$$\overline{x_1} - \overline{x_5} = T_{16} - T_{14}$$

$= 1.747 - 1.597$

$= 0.150$

Treatments are from same blocks; $SE(d_1) = 0.623$ from (3.17)

Let the level of significance, $\alpha = 5\%, df = 16$

$$LSD_\alpha = \left(t_{df,\alpha/2}\right)\left(SE(d_1)\right) \quad \textbf{(3.19)}$$

From t-table; t tabular = $t_{df,0.05} = t_{16,0.025} = 2.583$

$$LSD_\alpha = 2.583 \times 0.623$$

$= 1.609$

$\overline{x_1} - \overline{x_2} < LSD_{0.05}$ hence two treatment means are not significantly different.

Case 2: When treatments are in different blocks

For two treatments, take $T_{16} = \overline{x_1} = 1.747$ and $T_2 = \overline{x_2} = 1.682$

$$\overline{x_1} - \overline{x_2} = T_{16} - T_2$$

$$= 1.747 - 1.682$$

$$= 0.065$$

Treatments are from different blocks; $SE(d_2) = 0.586$ from (3.18)

Let the level of significance, $\alpha = 5\%$, $df = 16$

$$LSD_\alpha = \left(t_{df,\alpha/2}\right)\left(SE(d_2)\right) \quad \textbf{(3.20)}$$

From t-table; t tabular = $t_{df,0.05} = t_{16,0.025} = 2.583$

$$LSD_\alpha = 2.583 \times 0.586$$

$$= 1.514$$

$\overline{x_1} - \overline{x_2} < LSD_{0.05}$ hence two treatments are not significantly different i.e. T_{16} is not significantly better than T_2 at 5% level.

The best performing maize variety can be determined by comparing the highest yielding maize variety with the rest to find out whether there are some significant differences.

Table 27: Comparison of all treatment means with T16

Treat no.	$\overline{X_i}$	$\overline{X_i}$	$\overline{X_1} - \overline{X_i}$	LSD	Signif
T16	X1	1.747			
T2	X2	1.682	0.065	1.514	NS
T5	X3	1.629	0.118	1.514	NS
T12	X4	1.599	0.148	1.514	NS
T14	X5	1.597	0.150	1.609	NS
T19	X6	1.588	0.159	1.514	NS
T24	X7	1.587	0.160	1.514	NS
T23	X8	1.581	0.166	1.514	NS
T8	X9	1.580	0.167	1.609	NS
T22	X10	1.561	0.186	1.514	NS
T10	X11	1.441	0.306	1.514	NS
T11	X12	1.424	0.323	1.609	NS
T20	X13	1.411	0.336	1.514	NS
T9	X14	1.391	0.356	1.514	NS
T3	X15	1.333	0.414	1.514	NS
T18	X16	1.318	0.429	1.514	NS
T6	X17	1.290	0.457	1.514	NS
T17	X18	1.288	0.459	1.514	NS
T15	X19	1.275	0.472	1.514	NS
T7	X20	1.267	0.480	1.514	NS
T25	X21	1.248	0.499	1.514	NS
T21	X22	1.107	0.640	1.609	NS
T13	X23	0.844	0.903	1.514	NS
T4	X24	0.580	1.167	1.609	NS
T1	X25	0.359	1.388	1.514	NS

NS = Not significant
$*$ = Significant at 5% level

Mean differences are not significantly different at 5% level of significance. That is there is no significant difference between variety T16, the highest yielding and the other varieties. Therefore no variety is significantly better than the other.

CHAPTER 4

COMBINED ANALYSIS FOR THE TWO LOCATIONS

4.1 Introduction

An important objective of an on-farm or field research is often to examine which treatment is adapted to which kind of environment. A major reason for replicating experiments over multiple environments as locations or sites is to estimate the effects of treatments over a variety of environments.

The analysis of variance over different sites or seasons shows whether treatment effects change under different environmental conditions. For example, a maize breeder needs to know the area of adaptation of new maize varieties developed. To achieve this objective, the varieties are tested in field experiments repeated in several locations distributed in maize growing areas. The conclusions drawn from an experiment in a single locality will have little value for the whole, because performance of varieties will vary depending on the type of soil, amount of rainfall and rainfall pattern, and diseases and pests prevalent in different localities within the target area. When varieties respond in different ways to changes in environments we conclude that there is a variety by location interaction.

The purposes of multi-location tests of a set of treatments are:

i) To recognize if the area is reasonably homogeneous or if it should be divided into more homogeneous locations.

ii) To draw conclusions about the treatments themselves. This will enable us to recommend the use of particular treatment for location

iii) To recognize the superior treatments in maize variety experiment.

Most field experiments are conducted over two or more locations and years. Snedecor and Cochran (1967) described the procedure for combined analysis of one factor experiments, but do not describe the test of the average response of treatments over locations and years. The test of the main effect of locations or years may be of interest to researchers, but is not readily available in the literature. Much has not been done on combined analysis of multi locations on lattice design.

4.2 Combined analysis procedures

The following steps are followed when carrying out combined analysis for groups of experiments;

Step 1: Construct an outline of combined analysis of variance over locations

Step 2: Perform analysis of the locations separately.

Step 3: Test equality of experimental error variances or homogeneity.

For step 3, there are situations which depend on the number of l error mean square or variances, where l is the number of locations.

Case 1: When $l = 2$

F-test for testing the homogeneity of variance is applied. S_1^2 and S_2^2 are taken to be the mean square errors (MSE) for the two locations. The value of F statistics $\frac{S_1^2}{S_2^2}$ is tested against the tabulated F value at n_1 and n_2 degrees of freedom at 5% level of significance, where n_1 and n_2 degrees of freedom (df) of errors of the two locations respectively. The larger S^2 is the numerator value. If F computed is greater than F tabulated then the two locations are heterogeneous.

Case 1: When $l > 2$

Bartlett`s Chi-Square test is used when locations (populations) are more than two. It is designed to test equality of variances across locations against alternative that variances are unequal for at least two locations.

The null and alternative hypotheses are;

$$H_o : \sigma_1^2 = \sigma_2^2 = ... = \sigma_l^2 \quad \text{against}$$

H_1 : at least two $\sigma_i^2 's$ are not equal.

The statistic, $$\chi^2_{1-\alpha,l-1} = \frac{(N-l)\log_e S_p^2 - \sum_{i=1}^{l} n_i \log_e S_i^2}{1 + \frac{1}{3(l-1)}\left(\sum_{i=1}^{l}\frac{1}{n_i} - \frac{1}{(N-l)}\right)} \qquad \textbf{(4.1)}$$

Where; n_i = degree of freedom of the i^{th} location

S_i^2 = variance of the i^{th} location

N = total sample size $\left(N = \sum_{i=1}^{l} n_i\right)$

l = number of locations and S_P^2 = pooled variance.

The pooled variance is a weighted average of location variances and is defined as;

$$S_P^2 = \frac{\sum_{i=1}^{l}(n_1 - 1)S_i^2}{N-l} \qquad \textbf{(4.2)}$$

The $\chi^2_{1-\alpha,l-1}$ follows the chi-square distribution with $(l-1)$ degrees of freedom at α level of significance. If the calculated value of $\chi^2_{1-\alpha,l-1}$ is greater than tabulated value $\chi^2_{1-\alpha,l-1}$ then the null hypothesis of homogeneity of variance is rejected and the locations are significantly different at α level of significance. Locations are not homogeneous.

Step 4: Combined analysis is performed if the population is homogeneous.

4.2.1 Statistical model of a combined lattice design

The model of a combined lattice design is given by:

$$Y_{ijklm} = \mu + L_i + R(L)_{j(i)} + B_k + T_k + e_{ijklm}$$

Where Y_{ijklm} = the observed value

μ = the general mean yield.

L_i = Effect of the i^{th} location

$R(L)_{j(i)}$ = Effect of the j^{th} replication within the i^{th} location.

B_k = Effect of the k^{th} block

T_L = Effect of the l^{th} treatment

e_{ijklm} = Random error

4.2.2 General formats for combined analysis over multiple locations

Much has not been done on Anova formats of lattice design at multiple locations but there are two namely the standard and SAS formats given in tables 28 and 29 respectively.

Table 28: The standard Anova table format for lattice experiment at multiple locations

Source	df	Sum of Squares(SS)	Mean square(MS)
Locations (L)	$p-1$	SSL	MSL
Replicates within Locations	$p(r-1)$	SSR	MSR
Treatments (unadj.) (T)	$t-1$	SST	MST
Treatments (adj.)	$t-1$	SST_a	MST_a
Block(adj) (B)	$pr(k-1)$	SSB_a	MSB_a
Intra-block error	$p(k-1)(rk-1)$	SSE	MSE
Total	$prt-1$	SST_o	

Table 29: The SAS Anova table format for lattice experiment at multiple locations

Source	df	Sum of Squares(SS)	Mean square(MS)
Locations	$p-1$	SSL	MSL
Replications	$p(r-1)$	SSR	MSR
Blocks within Replication (adj)	$r(k-1)$	SSB	MSB
Component A	$r(k-1)$	SSB	MSB
Component B	$r(k-1)$	SSB	MSB
Treatments (unadj.)	k^2-1	SST	MST
Intra-block error	$(k-1)(rk-k-1)$	SSE	MSE
Randomized Complete Block Error	$p(k^2-1)(r-1)$	SSE_{RC}	MSE_{RC}
Total	$prt-1$	SST_0	

According to lattice procedures the blocks within replications sum of squares is further broken down into components A and B. If there is no repetition of the basic plan, the component B sum of squares is the same as the blocks within replications sum of squares. If there is repetition of the basic plan, the component A sum of squares reflects the variation among blocks that contain the same treatments.

4.3 Testing for homogeneity of experimental error variances (or populations)

Since there are two sets (populations) of data from Katumani and Kangundo, F-test is used to test for homogeneity i.e. a test on equality of variances is performed.

$$H_o : \sigma_1^2 = \sigma_2^2$$

Against

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

The test statistics is given by $F_{calc} = \frac{S_1^{\,2}}{S_2^{\,2}}$ where $S_1^{\,2} > S_2^{\,2}$

From table 17, mean square of intra block error for Kangundo $= S_1^{\,2} = 0.4300$

From table 8, mean square of intra block error for Katumani $= S_2^{\,2} = 0.05251$

$$F(16,16)_{calc} = \frac{S_1^{\,2}}{S_2^{\,2}} = \frac{0.4300}{0.05251}$$

$$= 8.19$$

$$F_{tabled} = qf(0.95,16,16) = 2.333$$

Since $F_{calculated} > F_{tabulated}$, H_o is rejected and therefore $\sigma_1^{\,2} \neq \sigma_2^{\,2}$

The two populations (locations) are significantly different at 5% level of significance and therefore combined analysis for the two locations cannot be done.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

The maize yield mean differences of all varieties in Kangundo were not significantly different at 5% level of significance as indicated in table 27. Variety T16 was the best in Kangundo with mean yield of 1.747 t/ha while variety T1 had the lowest mean yield of 0.359 t/ha. The best 10 varieties were; T16, T2, T5, T12, T14, T19, T24, T23, T8 and T22.

The best maize variety in Katumani was T9 with mean yield of 1.158 t/ha while T19 was the lowest with mean yield of 0.082 t/ ha. Treatments T4, T1, T6, T7, T11 and T19 were significantly different with best variety T9 and farmers should not plant them. The best 10 varieties from table 17 were T9, T22, T15, T23, T18, T10, T5, T2, T16 and T14. They treatment means were not significantly different from best variety T9.

The best variety T9 in Katumani did not do well in Kangundo as it was ranked 17^{th} position. The best variety T16 in Kangundo had a better yield in Katumani as it was ranked 9^{th} position. Farmers can be encouraged to plant it. The results revealed that varieties T22 and T23 did well in Katumani as well as in Kangundo. The same could be said of varieties T2 and T5.

I would recommend farmers in Kangundo to plant varieties T16, T2 and T5 while varieties T9, T22 and T15 should be planted in Katumani area. Since most of the varieties do not differ very much in yield capacity, more research should be done based on the diversity of the farmer's needs. Different types of field trials or experiments to investigate other factors like early maturity, dry matter content, susceptibility to pests, insects and diseases, fertilizer applications and so on should be conducted.

REFERENCES

Bose R.C and Nair K.R, Partially balanced incomplete block designs Sankya, 4,307-372 (1939).

Byerlee, D. and Eicher, C. K. (1997). *Africa's emerging maize revolution*. Lynne Rienner Publishers.

Cochran, W.G. and Cox, G.M. (1950), Experimental Designs, Wiley, New York

De Groote, H.,Ouma ,J.,Bett.,C.,Wekesa,E.,Odendo, M.,&Mose,L.(2004,February). Estimation of maize crop losses due to stem borers preliminary results of a national field survey in Kenya .*Proceedings of the 7th Eastern and Southern Africa Regional Maize Conference* (pp.401-406).Nairobi, Kenya :International maize and Wheat Improvement Centre (CIMMYT).

Eriyo, J (2013,3rd October) Prices of maize shoots up by 25pc as demand soars. Daily nation pp 24. *Proceedings of the 5th Africa Grain Trade Summit.*Mombasa.

FAO (2011). Fao statistical databases (faostat), agricultural data (http://faostat.fao.org). Fisher, R.A. 1926. The arrangements of field experiments. J. Ministry of Agriculture, England 33: 503-513.

Jayaraman, K. 1999. Forestry Research Support Programme for Asia and the Pacific. A statistical Manual for Forestry Research. Food and Agriculture Organisation (FAO) of the United Nations Regional Office for Asia and The Pacific Bangkok.

Idrees Nadia and Khan Muhammad Inayat, Design improvement using uniformity trials experimental data, Pak. J. Agri. Sci., Vol. 46(4), 2009.

Lentner, M. and Bishop, T. (1993) Experimental design and analysis (Second Edition).Valley Book Company, Blacksburg, Virginia

Ma, R.H. and J.B. Harrington. 1948. The standard errors of different designs of field experiments at the University of Saskatchewan. Scientific Agriculture 28: 461-473.

Makokha S, Kimani S, Mwangi W, Verkuijl and F Musembi (2010) Determinants of fertilizer and manure use in maize production in Kiambu district,Kenya. Mexico, D, CIMMYT and KARI

Miracle, M. P. (1965). The introduction and spread of maize in Africa. *The Journal of African History*,6(1):39–55.

Mohsen, A.A Abdel and Hegazy S.R.Abo ‘’Comparing the relative Efficiency of two experimental design in a wheat field trials’’

Morris, M. L. (1998a). Maize seed industries: A conceptual framework. In Morris, M. L., editor, *Maize Seed Industries in Developing Countries*, pages 35–54. Lynne Rienner Publishers, Inc., Boulder

Muhammad, L.,&Underwood,E.(2004).The maize agricultural context in Kenya .In A.Hilbeck &D.A.Andow(eds),*Environmental risk assessment of genetically modified orgasims:Vol.1A Case study of Bt Maize in Kenya* (pp.21-56).Cambridge,MA:CABI Publishing.

Mureithi, G (2005) Maize Varieties, Soil fertility improvement and appropriate agronomic practices The Soil Managent Project (SMP) paper.

NAFIS. National Farmers Information Sevices, www.nafis.go.ke

Patterson, H.D. and E.A. Hunter. 1983. The efficiency of incomplete block designs in national list and recommended list cereal variety trials. J. Agric.Sci., Camb. 101: 427-433.

Pingali, P.l. (2001).CIMMYT 1999-2000 world maize facts and trends .Meeting world maize needs: Technological opportunities and priorities for the public sector (CIMMYT Technical Report).Mexico, D.F: CIMMYT

Smale, M. and Jayne, T. (2003). Maize in Eastern and southern Africa: ”seeds” of success in retrospect eptd discussion paper no. 97. Technical report, IFPRI, Washington,

D.C. Smale, M., Byerlee, D., and Jayne, T. (2011). Maize revolutions in sub-saharan africa. (5659).

Snedecor G.W, and Cochran W.G (1980): Statistical methods; (7th edition). The IOWA State University Press. Ames, IOWA, U.S.A.

The LATTICE Procedure manual, Chapter 35

UNESCO Soil map of the world (1974):vol. VI Africa UNESCO; Paris. Pg 299Yates, F. 1936. A new method of arranging variety trials involving a large number of varieties. Journal of Agricultural Science 26: 424–455.

APPENDICES

Appendix 2: Maize yield data in tons per hectare for Katumani in the two replicates.

Katumani replication 1

Block 1	**Block 2**	**Block 3**	**Block 4**	**Block 5**
9 (1.190)	*21* (0.670)	*2* (0.920)	*3* (0.920)	*22* (0.610)
13 (0.600)	*8* (0.950)	*24* (0.660)	*10* (0.990)	*12* (0.560)
18 (0.890)	*14* (1.000)	*5* (0.865)	*25* (0.675)	*20* (0.580)
4 (0.395)	*11* (0.160)	*23* (1.160)	*6* (0.510)	*17* (0.610)
1 (0.390)	*16* (0.590)	*7* (0.520)	*15* (1.050)	*19* (0.045)

Katumani replication 2

Block 1	**Block 2**	**Block 3**	**Block 4**	**Block 5**
9 (1.140)	*13* (0.590)	*18* (0.930)	*4* (0.345)	*1* (0.350)
21 (0.875)	*8* (0.500)	*14* (0.620)	*11* (0.003)	*16* (1.055)
2 (0.750)	*24* (0.760)	*5* (0.915)	*23* (0.890)	*7* (0.080)
3 (0.360)	*10* (0.810)	*25* (0.830)	*6* (0.130)	*15* (1.100)
22 (1.590)	*12* (0.580)	*20* (0.740)	*17* (0.540)	*19* (0.090)

Key:
Italic numbers represent the different varieties.
Numbers in brackets represent the different variety yields

Appendix 2: Maize yield data in tons per hectare for Kangundo in the two replicates

Kangundo replication 1

Block 1	**Block 2**	**Block 3**	**Block 4**	**Block 5**
9 (1.460)	*21* (1.155)	*2* (2.240)	*3* (2.285)	*22* (0.480)
13 (0.335)	*8* (1.790)	*24* (2.050)	*10* (2.010)	*12* (1.360)
18 (1.850)	*14* (1.350)	*5* (1.430)	*25* (1.860)	*20* (0.970)
4 (0.710)	*11* (1.510)	*23* (1.810)	*6* (0.960)	*17* (1.420)
1 (2.750)	*16* (2.080)	*7* (1.205)	*15*(1.525)	*19* (1.054)

Kangundo replication 2

Block 1	**Block 2**	**Block 3**	**Block 4**	**Block 5**
9 (1.390)	13 (1.440)	18 (0.970)	4 (0.640)	1 (0.800)
21 (1.030)	8 (1.360)	14 (1.930)	11 (1.430)	16 (1.230)
2 (1.170)	24 (1.190)	5 (1.990)	23 (1.520)	7 (1.220)
3 (0.570)	10 (1.080)	25 (0.940)	6 (1.930)	15(1.060)
22 (2.265)	12 (1.480)	20 (1.590)	17 (0.900)	19 (1.590)

Key:
Italic numbers represent the different varieties.
Numbers represent the different variety yields

Appendix 3: R Procedure for EDA of data yields from Katumani

```
c=read.csv(file.choose())
> attach(c)
> require(graphics)
> c
   group   block  treatmnt          yield
1      1     1        9             1.190
2      1     1       13             0.600
3      1     1       18             0.890
4      1     1        4             0.395
5      1     1        1             0.390
6      1     2       21             0.670
.....................................
.....................................
.....................................
.....................................

47     2     5       16             1.055
48     2     5        7             0.08
49     2     5       15             1.100
50     2     5       19             0.090

> hist(yield,main='Histogram for
Katumani',xlab='Yield(Kg/ha)',ylab='Frequency',prob=TRUE)
> curve(dnorm(x, mean=mean(yield), sd=sd(yield)),type="l",
add=T)
> qqnorm(yield,main='Normal Q-Q Plot for
Katumani',xlab='Observed value',ylab='Expected normal value')
> qqline(yield,main='Normal Q-Q Plot for Katumani')
> boxplot(yield,main='Boxplot for Katumani')
> mean(yield)
> sd(yield)
```

Appendix 4: SAS Lattice procedure for analysing maize data yields from Katumani

```
data katumani;
input group block treatmnt yield@@;
cards;
      1     1           9     1.19
      1     1           13    0.6
      1     1           18    0.89
      ..........................
      ..........................
      ..........................
      ..........................

      2     5           7     0.08
      2     5           15    1.100
      2     5           19    0.09
      ;
proc lattice data=katumani;
var yield;
run;
```

Appendix 5: R Procedure for EDA of data yields from Kangundo

```
> f=read.csv(file.choose())
> attach(f)
> require(graphics)
> f
   reps block variety yield
1     1     1       9 1.460
2     1     1      13 0.335
3     1     1      18 1.850
4     1     1       4 1.710
.........................
.........................
.........................
.........................

47    2     5      16 1.230
48    2     5       7 1.220
49    2     5      15 1.060
50    2     5      19 1.590

> hist(yield,main='Histogram for
Kangundo',xlab='Yield(Kg/ha)',ylab='Frequency',prob=TRUE)
> curve(dnorm(x, mean=mean(yield), sd=sd(yield)),type="l",
add=T)
> qqnorm(yield,main='Normal Q-Q Plot for
Kangundo',xlab='Observed value',ylab='Expected normal value')
> qqline(yield,main='Normal Q-Q Plot for Kangundo')
>
> boxplot(yield,main='Boxplot for Kangundo')
>
> mean(yield)
```

Appendix 6: SAS Lattice procedure for analysing maize data yields from Kangundo

```
data kangundo;
input group block treatmnt yield@@;
cards;
      1     1     9     1.460
      1     1     13    0.335
      1     1     18    1.850
      ......................
      ......................
      ......................

      1     5     17    1.420
      1     5     19    1.054
      2     1     9     1.390
      2     1     21    1.030
      ......................
      ......................
      ......................

      2     5     7     1.22
      2     5     15    1.06
      2     5     19    1.59
;
proc lattice data=kangundo;
var yield;
run;
```