ADDIS ABABA UNIVERSITY

GRADUATE STUDIES PROGRAMME

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF STATISTICS

APPLICATION OF LINEAR MIXED MODEL TO INCOMPLETE BLOCK DESIGNS

By

Demeke Lakew

A thesis submitted to the Office of Graduate Studies of Addis Ababa University in partial fulfillment of the requirements for the Degree of Masters of Science in Statistics

June, 2010

ADDIS ABABA UNIVERSITY

GRADUATE STUDIES PROGRAMME

COLLEGE OF NATURAL SCIENCES

DEPARTMENT OF STATISTICS

APPLICATION OF LINEAR MIXED MODEL TO INCOMPLETE BLOCK DESIGNS

By

Demeke Lakew

Approved by the board of examiners:

______________________ ______________

Department Chairman Signature

______________________ ______________

Examiner Signature

Prof. Eshetu Wencheko ______________

Examiner Signature

Tel. 251-1-11-239465 **P.O.Box 1176 Addis Ababa, Ethiopia**

Dedication

I would like to dedicate this work to my mother and father:

Shegultie Belay

&

Lakew Workie

I pray to God carry on them jovial.

ACKNOWLEDGMENTS

First of all, I would like to give Glory to God for his goodwill in opening my eye to go through this way and helping me reach this success.

My heartfelt gratitude goes to Dr. Girma Taye, my advisor, who has been relentlessly giving me his invaluable advice and excellent scientific guidance in the best way of handling the research work that makes it finally come real.

I also would like to extend my sincere appreciation and thanks to Gondar University headquarters and staff members of Statistics Department who paved the way for me to win this chance. I also thank my instructors and all the staff of Statistics Department, AAU, for their wholehearted support for sharing and cooperation. A word of thank to my colleagues for their moral support and making the atmosphere favorable during the program and their stamina with one another at all time.

It is my deepest and warm gratitude to my father and mother Lakew Workie and Shegultie Belay for their encouragement and prayer and to all of my brothers and sisters who were always there to encourage and provided me with continuing motivation.

Last but not the least, I am grateful to my friend Bedilu Alamirie for financial and moral support to complete this work and all those who encouraged me in one way or the other.

Table of Content

List of Tables

List of Figures

LIST OF ACRONYMS

AIC: Akiake's Information Criteria

AICC: Akiake's Information Criteria Correction

ANOVA: Analysis of Variance

BIC: Schwarz's Bayesian Criteria

SLD: Simple Lattice Design

BLUE: Best Linear Unbiased Estimator

BLUP: Best Linear Unbiased Predictor

CAIC: Conditional Akiake's Information Criteria

CD: Cook's Distance

CS: Compound Symmetry

CSA: Central Statistical Agency

COVPARM: Covariance Parameter

COVRATIO: Covariance Ratio

COVTRACE: Covariance Trace

EBLUE: Empirically Best Linear Unbiased Estimator

EBLUP: Empirically Best Linear Unbiased Predictor

EIAR: Ethiopia Institute of Agricultural Research

GDP: Gross Domestic Product

GLM: General Linear Model

H: Hessian Matrix

HTML: Hypertext Markup Language

I: Identity matrix

IBD: Incomplete Block Design

IC: Information Criteria

ICC: Inter-class Correlation Coefficient

LME: Linear Mixed Effect

LN: Natural Logarithm

ML: Maximum Likelihood

MME: Mixed Model Equation

MSE: Mean Square Error

NASS: National Agricultural Statistical Services

NBRP: National Barley Research Program

NLME: Non Linear Mixed Effect

ODS: Output Delivery System

OLS: Ordinary Least Squares

PROC: Procedure

RCBD: Randomized Complete Block Design

REML: Restricted/Residual Maximum Likelihood

RLD: Restricted Likelihood Distance

SAS: Statistical Analysis Software

SPSS: Statistical Package for Social Sciences

SUSAN: Sub Saharan African Network

UN: Unstructured

US: United States

USAD: United States Agricultural Department

VC: Variance Component

ABSTRACT

The study was designed to examine the application of linear mixed model to incomplete block designs. In addition to this, it is planned to compare VARCOMP, ML and REML estimation methods for variance components of linear mixed models. Sixty three promising barley lines and one standard check cultivar which were obtained from EIAR have been evaluated for grain yield performance and adaptation across eight environments (combination of four locations by two fertilizers).

The mean grain yields for individual line ranged from 17.06 to 33.21 quintals per hectare and the mean grain yields for individual environment ranged from 16.80 to 44.214 quintals per hectare. The highest mean grain yield was observed at BEKOJI, while the lowest mean grain yield was registered at SHENO with both fertilizers doses (100 and 150 kg). When we compare each variety with specific environment, line 8 and 55, Variety 20 and 14, Variety 49 and 54 and Variety 48 and the Local check to be adapted and have best mean effects to BEKOJI, SHENO, HOLETTA and NORTH GONDAR respectively.

REML estimates for variance components are indistinguishable from classical techniques in case of balanced data. This implies optimal minimum variance properties and REML estimates do not rely on normality assumption. But, for unbalanced data, the REML estimation for variance components is different from classical estimates. We have seen from the differences that estimation of variance components benefits from REML but VARCOMP does not. Therefore, the REML approach is appropriate to estimate variance components in SLD and SLD with missing plots.

Varieties/lines listed above were recommended to release in Ethiopia that have similar ecologic zones of respective locations and REML techniques was recommended to be used for variance components estimates of linear mixed model for SLD and SLD with missing plots. It was found that diagnostics for Linear Mixed Model applied to estimate variance components are perhaps an area that needs exploration in the future.

CHAPTER ONE: INTRODUCTION

1.1. Background

The word agriculture is the English adaptation of Latin agricultura, from ager, "a field", [Latin Word Lookup] and cultura, "cultivation" in the strict sense of "tillage of the soil" [Latin World lookup]. Thus, a literal reading of the word yields "tillage of a field or of fields" ... and the study of agriculture is known as agricultural science. Agriculture is the production of food and goods through farming and forestry. It was the key development that led to the rise of human civilization; with the husbandry of domesticated animals and plants (i.e. crops) creating food surpluses that enabled the development of more densely populated and stratified societies. The major agricultural products can be broadly grouped into foods, fibers, fuels, and raw materials.

Cereal Crops in Ethiopia

Ethiopia is located between longitude 33°W and 48°E and between latitude 15.4°N and 3.4°S and it is an agrarian country with 85% of its population deriving their livelihood from small scale agriculture and 90% of the country's export based on agriculture (Environmental Policy, 1997). Agriculture also contributes about 50% of the country's gross domestic product (GDP) and supports around 70% of the raw material requirements of agro-industries (Ministry of Trade and Industry, 2008). This high proportion of the country's economic gains made from agriculture depends mainly on the existing diversity of indigenous crops or plants and livestock. Crop production is estimated to contribute on average about 60%, livestock 27%, and forestry and other subsectors around 13% of the total agricultural value (Ministry of Water Resource, 2001). Ethiopia is also one of the centers of diversity and origin of agricultural crop genetic resources. Thus, Ethiopia grows large varieties of crops which include cereals, pulses, oilseeds, stimulants, fibers, fruits, vegetables, root and tuber, and sugarcane. Cereals are the major food crops both in terms of the area they are planted to and volume of production obtained.

The number of cereals accessions[1] of Ethiopian origins that have been introduced to various international crop improvement program and seed companies is enormous: more than 1800 for wheat and more than 4500 for sorghum, around 2500 for barely and more than 900 and large numbers also reported for chickpea, lentil and finger millet (ICPPGR/FAO, 1997; Cavatassi *et al.*, 2006). They are grown in almost all the regions of Ethiopia and are produced in large volumes compared with other crops. Based on the report of Central Statistical Agency (CSA) in Meher Season of 2008/09 estimated that out of the total grain crop area, 78.23% (8.8 million hectares) was under cereals and that contributed 84.69% (about 144.96 million quintals) of the total grain production. The principal cereal crops are 'Teff', wheat, barley, which are primarily cool-weather crops grown predominantly at optimum altitude range from 1800 to 2200 meters and corn, sorghum, and millet are warm weather cereals cultivated mostly at lower altitudes along the country's western, southwestern, and eastern peripheries. As a result, Ethiopian Agriculture contributes the lion share of the gross domestic product, foreign currency earnings of the country from the sell of agricultural outputs abroad, as well as creates employment opportunity to the majority of the country's population. Hence, agriculture is the major sector expected to play a dominant role to bring about an overall sustainable economic growth to the country if strenuous efforts are made to modernize the farmer activity of the sector as a whole. Among a number of efforts that should be made by the concerned stakeholders to meet the desired goal mentioned above, the availability of reliable comprehensive and timely research on the performance of the researcher should know designs and modeling with corresponding estimation that leads to the desired goals.

Barley is the third most important cereal crop grown under rain fed conditions in Ethiopia. It is the predominant cereal in the major yield limiting factor in the Ethiopian barley high altitudes (above 2000 m a.s.l.; mostly between 2,000 and 3,500 meters) and cultivated in main season relies on June-September rainfall. It is used for the preparation of different food stuff, such as *injera*, porridge, *kolo*-eaten for breakfast and in between meals after pealed and roast it, and local drinks, such as *tella*, *kinito* (Hailu 2002).

1: Accessions are a sample of a crop variety collected at a specific location and time (Heywood, 1995).

According to a recent study, eating whole grain barley can regulate blood sugar (i.e. reduce blood glucose response to a meal) for up to 10 hours after consumption compared to wheat or even whole-grain wheat, which has a similar glycemic index. The effect is attributed to colonic fermentation of indigestible carbohydrates and also be used as a coffee substitute. These are not the only but also according to US recommendations for adult, barely nutritional value contains eight essential amino acids. As far as the straw of barley is concerned, it is one of the most important cereal straws produced in the mid-altitude and highland areas of Ethiopia, and is used for different purposes. The main use is as animal feed, especially during the dry season because traditionally cropping and livestock production have been closely integrated and complementary. Other uses of strew includes thatching, bedding materials, maintenance of soil fertility (being plowed back into the soil), and industrial uses (e. g. paper production, hardboard, egg trays, and for packaging in the glass industry) Capper et al (1988). Therefore, it is not surprising that barley occupies a significant proportion of the land cultivated annually in Ethiopia; according to the report of CSA 2008/09 (2001E.C.) the area cultivated under barley was 977,756.7 hectares and the total yield produced from those cultivated areas was 15,194,042 quintals.

1.2. Research Activities on Barley

Agricultural research for development of improved varieties, seed production and supply is focused on cereal crops. The maintenance of released varieties and the production of breeder and pre-basic seed will continue to be the duty and responsibility of the agricultural research institutions that have developed the varieties. The research institutions are expected to strengthen their breeding programs, variety maintenance, breeder and pre-basic seed production by acquiring more facilities and manpower. The improvement of suitable barley varieties for the different ecologic zones has world-wide importance and the same to Ethiopia intern. There is, therefore, a need for research interventions to develop improved varieties with higher yield, better resistance to lodging, tolerance to cold and drought, a higher nutritional value, and to strengthen the barley pathology research programs. In order for genetic resources to be efficiently utilized in plant-breeding programs firstly, it is necessary to determine whether useful genetic variation exists in the material and, secondly, to develop the most cost-effective method of introducing the potentially useful genes into commercially acceptable material (Kearsey, et al 1997). Barley-improvement programs, whether by breeding or direct gene manipulation, aim to match adaptation to the local environment and to enhance quality for processing (Ellis, et al 2002). Despite the long history of barley production in Ethiopia, the mean national barley yield is quite low. This low productivity can be ascribed to many factors, primarily those related to a low level of agricultural technology. The traditional farming community, accounting for over 95% of the cultivated land, relies on local varieties and crop production practices, and rarely utilizes improved crop and soil management techniques. EIAR's barley research activities, in close collaboration with other organizations, have included: (a) the development of varieties through breeding (b) the multiplication and distribution of breeder and basic seed (c) the coordination and execution of agronomic and crop management studies etc. The National Barley Research Program (NBRP) has general objectives of generating basic scientific information and applied technology to increase and sustain barley production in Ethiopia.

1.3. Motivations

Most of the time there are no particular specifications of model fitting and variance component estimation. Consequently, to a large extent the data will determine which estimation is appropriate. Before fitting any model, it is common to estimate covariance structure directly from the data. Often this gives some idea about which model might be appropriate. After one or more models are fitted, their quality can be judged by the covariance structure, and by comparing them with alternative specifications. These comparisons can be based on statistical significance tests or the use of particular model selections criteria.

One of the major problems in analysis of field trial is inappropriate accounting for field variability. Agricultural experimentation involves selection of experimental materials, selection of experimental units, planning of experiments, and collection of relevant information, analysis and interpretation of the result. There is no prefect experimental design nor are they any prefect analysis procedures known to account for all variations encountered in practice. Hence the quality of the result depends on how well the selected experimental design or method of analysis helps in estimation of variation (Girma 2005). Design based approach failed to account for such variation, especially, when block cannot be laid out successfully. As a result, experimental error mean square of classical analysis may be severely inflated (as sited by Girma 2005). High error mean square affects sensitivity of the test thus leading to inappropriate conclusions. The severely inflated experimental error mean square is the totality of both experimental and random effects. Therefore, random effect has its own contribution to the total variations. Hence, to minimize the severely inflated experimental error mean square we have to use a mixed effects model rather than fixed effect model.

Moreover, the analysis of incomplete block designs is different from the analysis of complete block designs in which comparison among variety effects and comparison among block effects are no longer orthogonal to each other. This requires improvement in both models in fitting and variance component estimation; therefore, in this thesis different methods of variance component estimation and model fitting to linear mixed models for incomplete block designs will be applied. It is hoped that this will lead to a proper model with best estimation of variance components for barley trials in order to draw a reliable conclusion.

1.4. Statement of the Problem

In a linear mixed-effects model, responses of a subject under experimental study are thought to be the sum of so-called fixed and random effects. In a mixed-effects model, random effects contribute only to the covariance structure of the data. The presence of random effects, however, often introduces correlations between cases. Though the fixed effect is the primary interest in most studies or experiments, it is necessary to adjust for the covariance structure of the data. The adjustment made in procedures like GLM-Univariate is often not appropriate because it assumes independence of the data.

The mixed effect model solves these problems by providing the tools necessary to estimate fixed and random effects in one model. Estimation in mixed effect model is based on maximum likelihood (ML) and restricted maximum likelihood (REML) methods, versus the analysis of variance (ANOVA) methods in VARCOMP. ANOVA methods produce an optimum estimator (minimum variance) for balanced designs, whereas ML and REML yield asymptotically efficient estimators for balanced and unbalanced designs. ML and REML thus present a clear advantage over ANOVA methods in modeling real data. The asymptotic normality of ML and REML estimators, furthermore, conveniently allows us to make inferences about the covariance parameters of the model, which is difficult to do in ANOVA. In addition to this, the use of mixed effects models represents a substantial difference from the traditional analysis of variance (ANOVA). Because the actual statistical approach is quite different in ANOVA and mixed effects models solve this and will lead to different results whenever we try to use different, and often more logical, covariance structures (Searle, et al 2006).

1.5. Objectives of the Study

1.5.1. General Objective

The general objective of this study is to compare different estimation methods of variance components of a linear mixed model for Lattice-design (in which it is incomplete block designs) and then select the best one and propose it for agricultural researchers for barely variety trials.

1.5.2. Specific Objectives

Based on the general objective of the study, the following are the specific objectives of this study.

- To assess the performance of linear mixed model for IBDs (Lattice-design).

- To compare different parameter estimation methods of variance components to barely trial data.
- To put forward appropriate estimation methods of variance components to linear mixed effects model to be used by agricultural researchers of barely trials.

1.6. Research Contribution

Several types of designs have been developed for agricultural field experiments. Using any of these designs may be possible for a particular project, but each design has its own advantages and disadvantages. The overriding principle for experimental design is: keep the design as simple as possible while satisfying the required level of scientific soundness. Researchers do not need a complex design with many experimental treatments, multi-factor interactions and difficult statistical analysis (like estimation of variance components) when a basic, simply designed experiment will produce the required information. Therefore, the outcome of the research will help agricultural researchers to conduct research with efficient use of limited research recourses and to determine optimum estimation methods for specific field crop trials. Moreover, the result helps researchers as a guideline for indicating possible sources of variation that might occur in research activities. On top of this, the result of this study will enable researchers to enhance the awareness of factors which influences the model. The result of this study will also be used as a source of information for researchers in the future. In general, the application of this research result will be expected to be beneficial for different researchers working in the area of agriculture and the result will be used as a basis for future study using linear mixed effects model in agricultural area.

1.7. Outline of the Thesis

In this work, interesting sub-problems will be identified and solved; the results of which can be useful in other contexts. The author has decided to structure the thesis according to the methodology used and thus the type of result. In Chapters 2, 3 and 4 related literature, analytical methods and model diagnostics are well explained. Brief discussions and conclusions emanated from the study are given under chapters 5 and 6. A more detailed description of the content of the chapters is as follows:

Chapter 2 provides a general review about linear mixed model with its advantages over ordinary regression and literature done specifically to variance components modeling. Chapter 3 describes estimating variance components based on VARCOMP, ML and REML with their details. Chapter 4 of the study demonstrates the details of model diagnostics like influential detection, model selection and test of normality assumption.

In Chapters 5, summary descriptive statistics, estimation of variance components and empirical linear unbiased estimates and predictions with tests and explanations are given. Finally, the overall conclusions and recommendations are provided in Chapter 6.

CHAPTER TWO: REVIEW OF THE LITERATURE

2.1. General Review

As cited by Girma, 2005 USAN (1996), the majority of research trials in Sub-Saharan Africa were planned as RCBD regardless of the size of experiments. Such designs are often applied at all levels of research including nursery, regional variety trials and national variety trials. This resulted in squeezing a large number of treatment combinations into a complete block which leads to loss of homogeneity within block. One of the basic principles in experimental design is that of reduction of experimental error. This can be achieved quite often through the device of blocking. This leads to designs such as randomized complete block designs or their family of incomplete block designs. A further reduction can sometimes be achieved by using blocks that contain fewer experimental units than varieties. It is because of that in the case of field plots experiments, the size of the plot is usually, though by no means always, fairly well determined by experimental and agronomic techniques. The experimenters usually aim toward a less block size than plots (Klaus and Kempthorne (2005)). Therefore, blocks the size of which is less than the number of varieties are called incomplete block designs.

Yates (1936) introduced the concept of balanced incomplete block designs and their analysis utilizing both intra- and inter-block information. Other incomplete block designs were also proposed by Yates (1940) who referred to these designs as lattice designs. After having introduced incomplete-block designs he called them 'incomplete block designs' in contrast with complete block designs. Incomplete-block designs are now widely used in plant breeding and variety testing around the world. At the beginning of their use, practically just after the Second World War, mainly lattice designs (square, rectangular and cubic) were used. Lattice design has been widely used in agricultural field experiments compared to other families of incomplete block designs. This is because Lattice design has the capacity to simultaneously handle two other sources of variation among experimental units. These are due to the row-blocking and the column-blocking. Because of such blocking, some estimates of the experimental error are removed. Moreover, Lattice design is suited to agricultural field trials where the gradient is parallel to both sides.

The use of mixed model has gone back to the beginning of 20 century. For example, Neyman et al (1935) have used linear mixed models in experimental design and is said to have used the term "variance components" in their paper. However, it was Daniel (1939) who firstly used the term variance components and he derived sampling variances for a three-way crossed classification random effects model with interactions and expected mean squares implying that the population effects for a random factor could be of finite size. Most of the historical cases covered above illustrate the application of linear mixed models to balanced data. For unbalanced data, Cochran (1939) applied the 1-way classification model using the analysis of variance method to estimate the variance components. Interestingly Ganguli (1941) and Crump (1946) further showed that the analysis of variance (ANOVA) method of estimating variance components can produce negative estimates for variances. By far the greatest advances in analyzing unbalanced data were made by Henderson (1953) in his work on the analysis of dairy cow records. Henderson presented three different ways of applying the ANOVA to variance component estimation. These three ways are known as Henderson's methods I, II and III.

Searle et al (1995) states that variance component estimation originated from estimating the error variance in the analysis of variance by equating the error mean square to its expected value. This procedure was then extended to random effects models for balanced data and then for unbalanced data. The beginning of variance components has revolved around the 1-way random effects classification model as: $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, where i= 1, 2, ..., a, the α_i ' s and ε_{ij} ' s are assumed to be random variables. In other words, the model focuses on a random selection from a population of all possible levels in the ANOVA model. μ is the mean, var(α_i) $=\sigma_\alpha^2$, var (ε_{ij}) $=\sigma_\varepsilon^2$ with all covariance equal to zero, j = 1, 2, ... n for observation of the data. Thus, the variances associated with the random effects are called variance components.

Searle (1988) used a "mixed model", even though it was not termed a mixed model as such. In describing a 2-factor-no-intraction situation, he concluded that one is a "measure of trial effect" and the other is a "measure of individual effect". This seems to be the first occurrence of the word "effect" which is ordinarily used in linear models these days. Searle also described his model as having one random and one non-random effect which looks like a clear specification of a mixed model, although not called such at the time. Searle et al (2006) also has given the following diagram or flow chart to help decide on the appropriate model to be fitted (as an adequate representation of the data):

"Is it reasonable to assume that levels of the factor come from a probability distribution?[2]

↓

No ↓ — Treat factors as fixed.

yes ↓ — Treat factors as random.

↓

Where does interest lie?

It is only in the distribution of the random effects.

↓

Estimate the variance of the random effects.

In both the distribution and the realized value of the random effect.

↓

Estimate the variance of the random effects and calculate best linear unbiased predictor of the realized values of the random effects."

2: This chart has been copied from Searle 2006.

Therefore, in recent years, mixed models have become invaluable tools in the analysis of experimental and observational data. In this thesis Lattice- design, which is incomplete block designs, is considered to recommend the best modeling in the corresponding design based on the appropriate estimation method.

2.2. Literature Related to Mixed Models

Ofversten (1993) has given a method of deriving exact tests for variance components in balanced mixed models. In particular he looked at a hypothesis of variance components of a model with one random factor. He stated that his proposed tests are unbiased and consistent under reasonable conditions. Reverter, et al (1996) used mixed model in assessing the efficiency of multiplicative mixed model to account for heterogeneous variance across herds in carcass scan traits from beef cattle at the University of New England. The result shows that the variance of the error terms of mixed model is much smaller than the variance of the error terms of fixed effect model. Chow and Shao (1988) also have undertaken research to obtain estimators for variance components which avoid non-positive estimates in random effects models. They use decision theoretic methods for estimating variance components. The estimators that are constructed are non-negative and some of them have smaller mean square errors than the classical estimators of ANOVA.

According to Romney et al. (2000) report the analysis that described how to determine (a) the influence of household (farm) and cow factors on milk yield, and (b) the relationships between milk yield and concentrate fed at different phases of lactation which were analyzed using mixed model effect in dairy production which in turn is an important source of income for many smallholder households in the highlands of East Africa.

Mora and Arnhold (2006) have examined genetic breeding values and variance components of popping expansion and grain production by means of a mixed linear model approach on 96 S_3 maize families. Best Linear Unbiased Predictors (BLUP) of family effect was obtained by considering the Restricted Maximum Likelihood (REML) method of variance component estimation. Family and residual variance component values were very similar in the Independence Chain Algorithm and the REML method. Heritability values showed imperceptible differences in the approximation between approaches. Differences in the standard deviation of these estimates were observed in the REML approach clearly showing the largest result. Heritability of grain production was moderate to high for popping expansion indicating that simple selection methods can be applied. Using an Independence Chain Algorithm and the BLUP approach for breeding values, no important changes were seen in family ranking, which was confirmed with high and significant Spearman's correlations values (rs) ranging from 0.9941%±0.004 to 0.9973%±0.001. Pearson's correlation between the BLUP values of popping expansion and grain production was low, negative and insignificant (rs=-0.320%±0.02). They concluded that an Independence Chain Algorithm could be an important tool to use in maize breeding like classical analysis using a mixed linear model procedure.

Yann et al (2007) compared forty-two paired organic and conventional winter wheat fields in three regions of western and central Germany; Leine Bergland, Soester Boerde, and Lahn-Dill Bergland. Based on the assumption that factors acting at various scales may affect biodiversity, they have compared such fields by using multiple spatial scales in order to understand how community richness is determined. They adopted a hierarchical approach to test the contribution of region, landscape heterogeneity, local management (organic vs. conventional) and location within field (edge vs. centre) to the species richness and abundance of spiders in cereals. Field pairs were located in areas ranging from structurally simple to structurally complex landscapes. In May and June 2003, spiders were sampled using pitfall traps. Linear mixed models were used to determine the relationship of spider diversity and abundance with regional spatial factors and landscape heterogeneity within a 500-m radius, as well as with local management and within-field location. Results they obtain within-field location of the traps and landscape heterogeneity were the best predictors of species richness. More species were found in field edges and in heterogeneous landscapes. Region and local management had no effect on species richness. Activity density was higher in field edges and differed among regions. Their main conclusions were the diversity of farmland spiders was influenced by differences at two of the spatial scales (edge vs. centre, simple vs. complex landscapes), but not at the two others (field management, region), emphasizing the importance of analyses at a variety of spatial scales for an adequate explanation of patterns in biodiversity. Their study suggests that promoting heterogeneity in land use at landscape scales is one of the keys to promoting spider diversity in agro-ecosystems.

Stephen (2007) studied a critical review of theoretical issues that underline the linear mixed effects (LME) and nonlinear mixed effects (NLME) models. These two areas were revisited under maximum likelihood and restricted maximum likelihood estimation frameworks. He also reviewed methods of estimating parameters in both linear and nonlinear mixed effects models. For NLME, he investigated the computational efficiency and accuracy of computational methods by approximating the log-likelihood function in nonlinear mixed effects models. He concluded by giving an insight into linear mixed effects models by analyzing a data set from livestock where he examined incorporation of random effects to study variations among rams (sires) and ewes (dams) and their influences on lamb weaning weight. Factors like year of birth of the lamb, sex of lamb, age at weaning, age of dam, ewe breed and ram breeds. His intention was to obtain heritability estimates which determine the proportion of the variation among offspring that have been handed down from parents out of these random estimates.

Osborne et al (2007) data set comprised of barley breeding lines and commercial varieties that were selected from a combined intermediate/ advanced stage breeding trial. The breeding lines represented a diverse range of genetic backgrounds although the program uses European malting varieties such as triumph for quality and dormancy traits. The study comprised a number of stages of testing that involve field testing to produce the grain samples followed by laboratory testing to measure the quality of the grain. A linear mixed model was adopted as the basis of the statistical analysis. All data were analyzed as a multiple environment trial, modeling the genetic variance or covariance matrix between environments. For grain data, the residual term was determined by variance between field plots as these data contained no additional duplication at the laboratory level. In contrast, the malting data was taken from a two-stage experiment where an extra level of design and randomization were employed at the micro-malting stage. Again the data were analyzed across multiple environments, but the model for analysis included a stratum variance for field plots as well as residual laboratory variation as the error term. On the basis of the results of Osborne et al, obtained using four commercial barley cultivars, the rheological parameters corresponding to the first derivative peak at 4.75 microseconds and trough at 100 microseconds were identified as potentially the most relevant for the prediction of the malting quality of barley and conclude that Hardness Index and rheological parameters recovered from the Crush-Response Profiles measured using grain data were shown to be influenced by genetic and environmental effects. These results indicate the use of rheological parameters may be a new tool for barley breeders to select for malt quality.

CHAPTER THREE: METHOD OF ANALYSES

3.1. The Study Area and Sources of Data

The data in this study were taken from yield of barley consisting of one commercially grown spring barley cultivar (Local check) and sixty three promising elite lines or varieties (Appendix A, Table 1). This experiment has been conducted at eight different environments in the central highlands of Ethiopia by the EIAR during the 2000E.C. cropping season. It was carried out by a Lattice Design with sixty four varieties in two replications. The sowing activities were undergone with uniform applications of inputs like seed rate, doses of fertilizer, and cultural practices used at each environment. Each location with a given fertilizer dose was considered as an individual environment. Lattice design has been widely used in agricultural field experiments compared to other family of designs and it is suited to agricultural field trials (Yates, 1940). That is why lattice design is used in this study among other family of designs.

The data were obtained from multi-location trial yield performance of sixty four barley lines or varieties evaluated over eight environments (four locations by two fertilizers combinations) having the same varieties in each replication for the purpose of comparing varieties at different location for environmental adaptation.

The study has FIVE explanatory variables (environment, rep(env), block(rep), variety and variety*env) and ONE response variable (yield measured in quintals per hectare). Environments that have been considered in this trial have the same potential variety verification. The major software employed for data analysis are SAS version 8.2, SAS 9.2 and an SPSS version 13.0 - constructing PP-plot, histogram and scatter plots for residual.

3.2. The Concept of Random and Fixed Factors

Mixed model methodology takes its name from the fact that the elements of the model underlying the analysis can be a mixture of fixed and random effects and the linear mixed models procedure expands the general linear model so that the data are permitted to exhibit correlated and non-constant variability. The linear mixed model, therefore, provides the flexibility of modeling not only by the means of the data but also by their variances and co-variances as well. The terminology mixed models is used when there are models for the fixed effects and one or more than one random effect. But, the generic term linear model is consistent for models where the fixed and random components are additive.

We need to describe and define concepts of balanced and unbalanced data and fixed and random effects that can be applied in mixed effects model.

A. Balanced and Unbalanced
   i. Data can be described as balanced when each experimental unit in the data set contains the same number of observations and as unbalanced when this is not the case.
   ii. Unbalanced data occur when the design of the experiment force the data to be so. Unbalanced data can also result from unfortunate circumstances or experimental carelessness, for example, if the experimenter loses some of the data points. As a result, the experimental units containing these missing observations vary in number of observations with respect to other experimental units.

B. Fixed and random effects Models
   i. A factor in a model is random if its levels consist of a random sample from a population of all possible levels. A model is termed a random effect model if all the factors in the variety structure are random.
   ii. A factor in a model is fixed if its levels are selected by a non-random process or consist of the entire population of all possible levels. A model is termed a fixed effects model if all the factors in the variety structure are fixed effects.

Therefore, in modeling, if the level of independent variable that occurs in the study is considered to be the only level of interest, each level has a fixed effect on the response variable and the independent variable is known as a fixed effect factor. Independent variables whose levels are determined or set by the experimenter are said to be fixed effect. However, if the levels are considered to be random sample from the population of possible levels, each level will have a random effect on the response variable and the variable is known as random effect factor. Random effects are classification effects where the levels of the effects are assumed to be randomly selected from infinite population of possible levels. A model which contains both fixed and random effect factors with error terms is called mixed model or mixed effects model. And if the relationship among these factors, the error term and the response variable is linear we call such a model as linear mixed model. Hence, mixed-effects model is a generalization of the standard linear model that enables the analysis of data generated from several sources of variation instead of just one (As Cited in SAS, 1996). Mixed model is a further extension that permits random effects as well as fixed effects in the linear predictor.

The only source of randomness in a fixed effect models is the error term and the parameter estimation could be done using Ordinary Least Square (OLS) technique. But, simple ordinary least square criteria could not be employed in models containing two independent random variables. Thus, the estimation of the mixed effect model is based on the least squares method and the variance components corresponding to different sources of randomness. These are estimated by maximum likelihood or restricted maximum likelihood criteria accordingly (Patterson and Thompson, 1971) to get estimates and/or predictors for both fixed and/or random factors.

As Ethiopia is the center of diversified barley species considering all species is tedious. This leads us to use variety as a random effect. Variety*environment is also considered as random effect since variety is random too. Environment is considered as fixed effect since it is predetermined by the experimenter as they are interested to compare the mean product effect of barley in each environment and number of replications and blocks that they considered are fixed on their interest rep(env) and block(rep) also fixed effects.

3.3. The model

3.3.1. Description of the Mixed Model

Here, we will introduce linear mixed model and its properties relevant to this thesis. The starting point is the traditional fixed effects linear model written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{.......} 1$$

,where **y** is an N **x 1** vector of response variable, β is a p **x** 1 vector of fixed effects parameters, **X** is a known N **x** p coefficient matrix and $\boldsymbol{\varepsilon}$ is an error vector defined as:

$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{E(y)} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ and thus has $E(\boldsymbol{\varepsilon}) = \mathbf{0}$. And the dispersion matrix var($\boldsymbol{\varepsilon}$) = $\sigma_{\varepsilon}^2 \boldsymbol{I}_N$. **X** is often a matrix of zero's and one's, known as *design* matrix.

In mixed models the random effects of the model can be represented as **Zu,** of a nature that parallels $\mathbf{X}\boldsymbol{\beta}$. **u** will be the vector of the random effects that occur in the data and **Z** the corresponding design matrix, usually design matrix. Moreover, **u** can be partitioned into a series of **r** sub-vectors:

$$\mathbf{u} = [\boldsymbol{u}_1^{'} \, \boldsymbol{u}_2^{'} \ldots \boldsymbol{u}_r^{'}] \quad \text{.......} 2$$

where each sub-vector is a vector of effects representing all levels of a single factor occurring in the data, be it a main effects factor, an interaction factor or a nested factor and **r** represents the number of such random factors.

Incorporating **u** into (2) gives a general form of model equation for a linear mixed model as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad \text{.......} 3$$

with **β** representing fixed effects and **u** being for random effects. **X** and **Z** are the corresponding model matrices, with **Z** often a design matrix, and $\boldsymbol{\varepsilon}$ is a vector of residual errors. It is

$$\mathbf{E(y)} = \mathbf{X}\boldsymbol{\beta} \text{ and } \mathbf{E(y \mid u)} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \quad \text{.......} 4$$

As a result

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{E}(y|\,u) \quad \text{.......... 5}$$

E(y|u) is the conditional mean of **y**, given that **u** represents the actual random effects as they occur in the data. **E(y|u)** we would mean **E(Y|U =u)**= $\mathbf{X}\boldsymbol{\beta} + \mathbf{ZU}$ where **Y** and **U** would be vectors of random variables for which **y** and **u** are the realizations in the data. Thus **E(Y|U =u)**= $\mathbf{X}\boldsymbol{\beta} + \mathbf{ZU}$ would be the expected value of the random variable **Y**, given that the random variable **U** has the value **u**.

For (3) we therefore have

$$\mathbf{E(y)} = \mathbf{X}\boldsymbol{\beta} \ \text{ and } \mathbf{E}\,(\boldsymbol{\varepsilon}) = \mathbf{0} \quad \text{.......... 6}$$

To $\boldsymbol{\varepsilon}$ we now attribute the usual variance-covariance structure for error terms: every element of $\boldsymbol{\varepsilon}$ has variance σ_ε^2 and every pair of elements has covariance zero, i.e.

$$\mathbf{Var}(\boldsymbol{\varepsilon}) = \mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}_N \quad \text{.......... 7}$$

Similar properties are attributed to the elements of each $\boldsymbol{u_i}$:

$$\mathbf{Var}(\mathbf{u}_i) = \sigma_i^2 \mathbf{I}_{qi} \ \forall_i \quad \text{.......... 8}$$

with q_i being the number of elements in **u**, i.e., the number of levels of the factor corresponding to $\mathbf{u}_i$ that are represented in the data.

$$\mathrm{cov}(u_i, u_j') = \mathbf{0}, \forall\, i \neq j; \quad \text{.......... 9}$$

and similarly for all elements of U and $\boldsymbol{\varepsilon}$:

$$\mathbf{cov}(\mathrm{U}, \varepsilon') = \mathbf{0} \quad \text{.......... 10}$$

Utilizing (8)-(10), the variance-covariance structure of u is:

$$G = \mathrm{var}\,(\mathrm{U}) = \begin{bmatrix} \sigma_1^2 I_{q1} & . & . & . \\ . & \sigma_2^2 I_{q2} & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & \sigma_r^2 I_{qr} \end{bmatrix} \quad \text{.......... 11}$$

where each $\{\sigma_i^2 I_{qi}\}$ is a diagonal matrix of dimension q_i, i= 1, 2, 3, ..., r.

Then partitioning Z conformably with U of (2) as Z = [Z, Z, Z, ... , Z_r] gives

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\,\mathbf{U} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^{r} Z_i U_i + \boldsymbol{\varepsilon} \quad \text{......} 12$$

Hence from (3) to (11)

$$\mathbf{V} = \mathbf{var}\,(\mathbf{y}) = \mathbf{ZGZ'} + \sigma_\varepsilon^2\mathbf{I} = \mathbf{ZGZ'} + \mathbf{R} = \sum_{i=1}^{r} \sigma_i^2\, \mathbf{Z}_i\mathbf{Z}_i' + \sigma_\varepsilon^2\mathbf{I}_N \quad \text{......} 13$$

There are several advantages to the application of the mixed model to agricultural data. Duchateau et al (1998) gives the specific features of the mixed model as advantages that include:

- Complex data structures can be described in a natural way by the mixed model;
- The mixed model framework allows a flexible choice of the appropriate inference space;
- A mixed model also allows the prediction of random effects of interest by best linear unbiased prediction (BLUP) etc.

The mixed linear model represented by $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\,\mathbf{u} + \boldsymbol{\varepsilon}$ is the model presented by Henderson (1975), and studied extensively by him and his students and colleagues in genetics applications. It is now the standard form of the mixed model referenced in most statistical research and implemented in computer programs.

3.3.2. Description of the Application of the Model Used in this Study

In the present study we consider the following:

- Y is a 1024x1 vectors of yield of barley data obtained from sixty four different varieties and eight environments with the application of two replications to each using 64x8x2 = 1024 as described in Section 3.1.

- X is a 1024x41 design matrices relating the fixed effect (intercept, env, rep(env) and block(rep)) that contains 0's and 1's. The design matrix X is described as:

  X_{i1} = intercept of the model

  $X_{i2} = \begin{cases} 1, & \textit{if data is from environment } 1 \\ 0, & \textit{otherwise} \end{cases}$

  $X_{i3} = \begin{cases} 1, & \textit{if data is from environment } 2 \\ 0, & \textit{otherwise} \end{cases}$

  .

  .

  .

  $X_{i9} = \begin{cases} 1, & \textit{if data is from environment } 8 \\ 0, & \textit{otherwise} \end{cases}$

  $X_{i10} = \begin{cases} 1, & \textit{if data is form } \text{rep } 1 \text{ at } \textit{environment } 1 \\ 0, & \text{otherwise} \end{cases}$

  $X_{i11} = \begin{cases} 1, & \textit{if data is form } \text{rep } 2 \text{ at } \textit{environment } 1 \\ 0, & \text{otherwise} \end{cases}$

  .

  .

  .

  $X_{i22} = \begin{cases} 1, & \textit{if data is form } \text{rep } 1 \text{ at } \textit{environment } 8 \\ 0, & \text{otherwise} \end{cases}$

  $X_{i23} = \begin{cases} 1, & \textit{if data is form } \text{rep } 2 \text{ at } \textit{environment } 8 \\ 0, & \text{otherwise} \end{cases}$

  $X_{i24} = \begin{cases} 1, & \textit{if data is form } \text{rep } 1 \text{ at } \textit{block}1 \\ 0, & \text{otherwise} \end{cases}$

  $X_{i25} = \begin{cases} 1, & \textit{if data is form } \text{rep } 1 \text{ at } \textit{block}2 \\ 0, & \text{otherwise} \end{cases}$

  .

  .

  .

  $X_{i40} = \begin{cases} 1, & \textit{if data is form } \text{rep } 2 \text{ at } \textit{block}7 \\ 0, & \text{otherwise} \end{cases}$

  $X_{i41} = \begin{cases} 1, & \textit{if data is form } \text{rep } 2 \text{ at } \textit{block}8 \\ 0, & \text{otherwise} \end{cases}$, where i = 1, 2, 3, ..., 1024.

- β is a 41x1 vector of fixed effect parameters.

- Z is a 1024x576 design matrices relating random effects (variety and variety*environment). The design matrix of Z is in the form:

  $Z_{i1} = \begin{cases} 1, & if\ data\ is\ for \text{ BSP SPS 46/95} \\ 0, & \text{otherwise} \end{cases}$

  $Z_{i2} = \begin{cases} 1, & if\ data\ is\ for \text{ BYDV 34/97} \\ 0, & \text{otherwise} \end{cases}$

  $Z_{i3} = \begin{cases} 1, & if\ data\ is\ for \text{ BYDV 42/97} \\ 0, & \text{otherwise} \end{cases}$

  .

  .

  .

  $Z_{i64} = \begin{cases} 1, & if\ data\ is\ form\ \ local\ ck \\ 0, & \text{otherwise} \end{cases}$

  $Z_{i65} = \begin{cases} 1, & if\ data\ is\ form \text{ BSP SPS 46/95 at } environment\ 1 \\ 0, & \text{otherwise} \end{cases}$

  .

  .

  .

  $Z_{i576} = \begin{cases} 1, & if\ data\ is\ form\ \ local\ ck\ \ at\ environment\ 8 \\ 0, & \text{otherwise} \end{cases}$, where i = 1, 2, 3,..., 1024.

- **U** is a 576x1 vector of unobservable random effects variables (U can be partitioned as U' = $[U_1^{'}, U_2^{'}]$; variety and variety*environment in this study). The distribution of U is considered to be normal with mean vector zero and variance-covariance matrix G.
- $\boldsymbol{\varepsilon}$ is a 1024x1unobservable vector of random residuals. The distribution of $\boldsymbol{\varepsilon}$ is a 1024 dimensional normal with mean vector zero and variance-covariance matrix R.
- The vectors U and $\boldsymbol{\varepsilon}$ are statistically independent, i. e.; cov (U, ε) = 0.
- The fixed components of the model and the error terms are uncorrelated, i. e.; cov(X, ε) = 0.
- U and ϵ are normally distributed with: $E\begin{bmatrix} U \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $var\begin{bmatrix} U \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$.

It was described in Section 3.2 that the study has two random variable for the data set i. e. ; variety and variety*environment. Therefore, it is a mixed model with two random variables and three fixed effects (env, rep(env), block(rep)) to analyze. The application of the theory of the general mixed model to the data set will lead to the following: **Var (U)** = $\sigma_u^2 I_{576}$;

$$Y = X\beta + ZU + \varepsilon = X\beta + \sum_{i=1}^{r} Z_i U_i + \varepsilon = X\beta + Z_1 U_1 + Z_2 U_2 + \varepsilon$$

$$V = var(y) = ZGZ' + \sigma_\varepsilon^2 I_{1024} = \sum_{i=1}^{r} \sigma_i^2 Z_i Z_i' + \sigma_\varepsilon^2 I_{1024}, \; r = 1, 2.$$

3.4. Data Analysis

In analyzing data via a linear mixed model, we are faced with the determination of variance-covariance structure. Thus, let us make a distinction between fixed effects that determine the level (expected mean) of observations, and random effects that determine variance. For every model at least there exist one fixed effect (mean) and one random effect (residual variance). Since variety and variety*environment are random factors in the dataset, there exists variance component for the respective random factors in addition to the residual variance. In this study there are two components contributing to the total variance of the observations: variety and variety*environment as well as a residual variance component. In predicting variety and variety*environment that gives best variety adaptation, we use best Linear Unbiased Prediction (BLUP).

3.4.1. Methods Used for Estimating Variance Components of Parameters

The model fitting consists of three parts; estimating variance parameters, fixed effects, and random effects. Depending on the data structure and the circumstances during measuring, estimation can be based on sufficiently large number of experimental observations or field recorded data. In general, we have to estimate variances if:

- ✓ We are interested in a new trial, from which on parameters are available;
- ✓ Variances and covariance might have changed over time;
- ✓ Considerable changes have occurred in a population, example, due to recent factors.

Therefore, we can undertake variance component estimation if there are sufficient reasons for regular estimation of variance-covariance components. There is an intensive review and discussion of theoretical aspects and application of estimation methods used in mixed effect models (Littell, 1996). Different from GLM, whose method of estimation is based on the ordinary least squares (OLS) estimation, the mixed model is based on different principles for estimation. As a result, mixed model has several options for the method of estimation including REML (Restricted/Residual Maximum likelihood), ML (Maximum Likelihood), MIVQUE0 (Minimum Variance Quadratic Unbiased Estimation), and Types I, II, and III (specifications apply only to variance component models with no subject =effects and no repeated statement). ML and REML are based on the maximum likelihood estimation approach which requires the assumption that the distribution of the dependent variable (error term and the random effects) is normal while MIVQUE0 does not require normality assumption for computing the estimators. However, according to Ngo, L and Rand, R. (2002) the evidence from simulation studies suggested favoring ML and REML over MIVQUE0. In addition to the methods of estimation, the mixed model also allows specification of a correlation or covariance structure with the TYPE option. So, there are a number of methods used to estimate variance-covariance components. This study explores some of these.

3.4.1.1. Henderson methods

There are three methods of Henderson which include method I, method II and method III. It should be noted that method I cannot be applied to mixed models and hence it has been left out. Also method II is not very practical in estimating variance components and some of its disadvantages include:

- Negative estimates can arise.
- Sampling variances of estimators are not obtainable in closed forms except under certain conditions.

Method III

The reasons that consider method III here is because Method III is used by a SAS procedure VARCOMP which calculates ANOVA (via method III) estimators from the SAS type I sums of squares.

Method III is based on borrowing sums of squares from the analysis of fixed effects models. The sums of squares used are the reductions in sums of squares due to fitting one model and various sub-models of it. We therefore begin a description of Method III with a brief summary of these sums of squares.

The following theory has been taken from Searle et al (2006, pp. 202-205). In writing a general mixed model equation as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon}$ we clearly distinguish between fixed effects and random effects, representing them by $\boldsymbol{\beta}$ and $\mathbf{u}$, respectively. Suppose for the moment that we remove this distinction and combine $\boldsymbol{\beta}$ and $\mathbf{u}$ into a single vector $\mathbf{b}$ and write the model equation as $\mathbf{Y} = \mathbf{Wb} + \mathbf{e}$ 14

The model in $\mathbf{Y} = \mathbf{Wb} + \mathbf{e}$ can be considered to be a fixed effects model forgetting all about the differences between fixed and random effects (i.e., $\mathbf{var}\ (\mathbf{u}) = \mathbf{0}$). Hence our best linear estimator of $\mathbf{Wb}$ is given as BLUE $(\mathbf{Wb}) = \mathbf{Wb^0} = \mathbf{W(W'W)^{-}W'y} = \mathbf{WW^{+}y}$, where, $\mathbf{W}^{+} = (\mathbf{W}'\mathbf{W})^{-}W'$ our residual sum of squares is now (after fitting the model):

$$\mathbf{SSE} = (\mathbf{y} - \mathbf{Wb^0})'(\mathbf{y} - Wb^0) = \mathbf{y'y} - \mathbf{y'}Wb^0 - (Wb^0)'y + (Wb^0)'(Wb^0) = y'y - y'WW^{+}y$$

So that $\mathbf{y'y} = \mathbf{SSE} + \mathbf{y'}WW^{+}y$

$\mathbf{SST} = \mathbf{SSE} + \mathbf{SSR}$

To estimate the variance components of the random effects, partition $\mathbf{Wb}$ in the following way.

$\mathbf{E(y)} = \mathbf{W_1b_1} + \mathbf{W_2b_2}$ with the following results that $\mathbf{R(b)} = \mathbf{R(b_1, b_2)}$

$$\mathbf{R(b_1, b_2)} = \mathbf{y'}\ [\mathbf{W_1\ W_2}][\mathbf{W_1\ W_2}]^{+}\mathbf{y}$$

So that the difference between the two reductions SSR is given as

$$\mathbf{R(b_2 \mid b_1)} = \mathbf{R(b_1, b_2)} - \mathbf{R(b_1)} = \mathbf{y'}\ [\mathbf{W_1\ W_2}][\mathbf{W_1\ W_2}]^{+} - \mathbf{y'}\ \mathbf{W_1}W_1^{+}\mathbf{y}$$

(Note that $\mathbf{WW^{+}} = \mathbf{W(\ W'W)^{-}\ W}$)

Therefore $\mathbf{R(b_2 \mid b_1)} = \mathbf{y'\ M_1\ W_2}(W_2'\mathbf{M_1\ W_2})^{-}\ W_2'\mathbf{M_1y}$, where, $\mathbf{M_1 = I - W_1(W_1'W_1)^{-}W_1' = M_1' = M_1^2}$ with $\mathbf{M_1W_1 = 0}$. 15

Keeping in mind that $\mathbf{R(b_1 \mid b_2) = y'\ M_1\ W_2(W_2'M_1\ W_2)^+\ W_2^-\ WM_1y}$ and $\mathbf{M_1}$ is given as above we see that $\mathbf{E[R(b_1 \mid b_2)] = tr\{\begin{bmatrix} W_1 \\ W_2 \end{bmatrix}[M_1W_2(W_2M_1W_2)]^{-}W_2'M_1\}}$.

Analogously for the mixed model of this thesis have:

$$\mathbf{ER\ (u \mid \beta) = tr\ [Z'MZE\ (uu')] + \sigma_\varepsilon^2(r[X\ Z] - r[X]) = tr(Z'MZ\{_d\sigma_i^2 1_{qi}\} + \sigma_\varepsilon^2(r[X\ Z] - r[X])}$$

$$\mathbf{= \textstyle\sum_i \mathit{tr}(Z_i'MZ_i)\sigma_i^2 + \sigma_\varepsilon^2(r[X\ Z] - r[X])} \quad \text{.........16}$$

The mixed model is given as in equation $\mathbf{y = X\beta + ZU + \varepsilon}$ with rank of X = r(X) = r and u representing the random levels as stated earlier, containing t levels for just two random factors with variance $\boldsymbol{\sigma_u^2}$. As a result **Z** has full column rank, t, with its columns summing to 1. Hence we have the following relationship of the columns of **Z** to **X**: r[**XZ**] = r(X) + t - 1 = r+ t – 1. Then it gives the following estimation using Henderson's method III:

SSE = sums of squares of the error terms = $\mathbf{y'y - R(b, u)}$

$\mathbf{R(u \mid b) = R(u, b) - R(b)}$ with $\mathbf{E(SSE) = \{N - (r{+}t - 1)\}}\ \sigma_\varepsilon^2$ and

$\mathbf{E[R(u \mid b)] = \sigma_u^2\ tr(Z'Z - ZX(X'X)^{-}X'Z) + \sigma_\varepsilon^2\ [r(X) + t - 1 - r(X)]}$

Hence our estimates of $\boldsymbol{\sigma_\varepsilon^2}$ and $\boldsymbol{\sigma_u^2}$ using the Henderson method III are:

$$\widehat{\sigma}_\varepsilon^2 = \frac{\mathbf{y'y - R(b)}}{\mathbf{N - r(X) - t + 1}} = \frac{SSE}{\mathbf{N - r(X) - t + 1}} \quad \text{.........(17a)}$$

and

$$\widehat{\sigma}_u^2 = \frac{\mathbf{R(u \mid b) - \widehat{\sigma}_\varepsilon^2\ (t-1)}}{\mathbf{tr(Z'Z - ZX(X'X)^{-}X'Z)}} = \frac{\mathbf{SSE - \widehat{\sigma}_\varepsilon^2\ (t-1)}}{\mathbf{tr(Z'Z - ZX(X'X)^{-}X'Z)}} \quad \text{.........(17b)}$$

where $(X'X)^-$ is a generalized inverse of (X'X) and *tr* is the trace operator.

Our link to the next section would be to look at fitting a mixed model equations with their (BLUP) solutions because Searle et al (2006) states that an interesting aspect of the mixed model equation is that the elements in them can be used to set up iterative procedures for calculating solutions to the Maximum Likelihood (ML) and the Restricted Maximum likelihood (REML) equations for estimating variance components. Furthermore the same elements can also be used for setting up and calculating the information matrices of ML and REML. Therefore we consider model fitting before to see ML and REML.

3.4.1.2. Henderson's Mixed Model Equations

The reason for considering Henderson's mixed model equations and their BLUP solutions is because the algorithms and methods of variance component estimation (such as ML and REML to be used later on) used by statistical packages such as SAS (which will be used extensively for this thesis) are based on the mixed model equations attributed to Henderson (1953). The algorithm uses BLUP estimation of the random effects (variety and variety*environment). A good general summary of the derivation of the mixed model equations barley trial data given above with all given assumptions is as follows.

They are derived by maximizing the joint probability density function of y and u, for which assume that all levels of **u** pertain to the same source of variation, the **var(ε) = R** and **var (u) = G** of order **q** and **cov(u, ε) = 0**.

$$\mathbf{f(y,u) = f(y \mid u).f(u)} = \frac{exp\{-\frac{1}{2}[y - X\beta - ZU)' R^{-1}[y - X\beta - ZU] + U' G^{-1}U\}}{(2\pi)^{0.5(N+q)} \mid R \mid^{\frac{1}{2}} \mid G \mid^{\frac{1}{2}}}$$

If we take the partial derivatives of f(y, u) and equate them to zero with respect to elements first of β and then of u gives, using $\widetilde{\beta}$ and ũ to denote the solutions

$$X'R^{-1}X\widetilde{\beta} + X'R^{-1}Z\widetilde{u} = X'R^{-1}y \text{ and}$$

$$X'R^{-1}X\widetilde{\beta} + (Z'R^{-1}Z + G^{-1}) = Z'R^{-1}y$$

We can also rewrite these equations as:

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ X'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \widetilde{\beta} \\ \widetilde{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \quad \text{.........} 18$$

where $\mathbf{R} = \sigma_\varepsilon^2\mathbf{I}$ and observing the fixed effects reduces (18) to:

$\mathbf{Z'M\tilde{u} = Z'My}$, **with** $\mathbf{M = I - X'(X'X)^{-} X'}$ and estimates of variance components are as (17),

$$\hat{\sigma}_\varepsilon^2 = \frac{y'y - \tilde{u}Z'y - \tilde{\beta}X'y}{N - r(X) - t + 1} \qquad \text{17a')}$$

and

$$\hat{\sigma}_u^2 = \frac{\tilde{u}Z'My - \hat{\sigma}_\varepsilon^2\,(t-1)}{tr(Z'MZ)} \qquad \text{(17b')}$$

Searle et al (2006, pp. 276-277) gives the BLUP solutions as:

$$X'V^{-1}X\tilde{\beta} = X'V^{-1}y \text{ and } \tilde{u} = G^{-1}V^{-1}(y - X\tilde{\beta}). \qquad 19$$

3.4.1.3. Maximum Likelihood (ML)

Searle et al (2006) state that the method of maximum likelihood estimation was developed by Fisher in 1925 and then applied by Crump in 1951. ML estimation is an old, yet clear, conceptually very simple, and efficient way of estimating the variance components of data. The crucial requirement in estimating the variance components of a set of data using ML technique is the assumption of underlying probability distributions for the data. For a given model of analysis, parameters to be estimated and data with a specified distribution, we can calculate the likelihood of particular numeric values of the parameters, i.e. how likely it is that the data have been sampled from a population with these parameter values. This is analogous to probability calculations where we determine the probability of observing a specific set of data for given parameter values, but with 'cause' and 'effect' reversed. ML estimates are then, by definition, the parameter values for which the likelihood is maximized.

As reviewed by Harville (1977), ML estimators are consistent, asymptotically normal and efficient, i.e.; all information available is utilized in an optimal way. Moreover, they are well defined for cases which cannot be accommodated by standard ANOVA models.

3.4.1.3.1. The Maximum Likelihood Equations

In general, if a vector X of random variables is distributed- $N(\mu, V)$ we know that the density function of X is given as $\mathbf{f(x)} = \frac{\exp\{-0.5(x-\mu)'V^{-1}(x-\mu)\}}{(2\pi)^{0.5N}|V|^{0.5}}$.

So for our **y** we have $\mathbf{y} \sim N_N(X\boldsymbol{\beta}, V)$ is a function of multivariate normal distribution with parameters in $\boldsymbol{\beta}$ and **V** to be $\mathbf{f(y)} = \frac{\exp\{-0.5(y-X\beta)'V^{-1}(y-X\beta)\}}{(2\pi)^{0.5N}|V|^{0.5}}$ 20

where N is the length of **y** and $|V|$ is the determinant of V. The function gives the probability of finding a certain y given the parameters. The parameters are the means in $X\boldsymbol{\beta}$ ("location parameters") and the variances in **V** ("dispersion parameters"). However, this function can also be used the other way round i. e.; if we have observed data, it gives us the probability of having such data for certain parameter values. When the data y is known, **f(y)** is a likelihood function and this function can be maximized in the parameters, i. e.; we want to find the parameters for which **f(y)** has highest value and instead of maximizing **f(y)** we now maximize $\mathbf{L} = \mathbf{L}(\boldsymbol{\beta}, V | \mathbf{X}, \mathbf{y})$, which is the log likelihood function, in equation (20) by taking the log of both sides of the equation and maximize log L with respect to $\boldsymbol{\beta}$ and σ_i^2 as $\boldsymbol{\ell}$ to gives us:

$$\boldsymbol{\ell} = \log L = -\frac{1}{2}N\log 2\pi - \frac{1}{2}\log|V| - \frac{1}{2}(y-X\beta)'V^{-1}(y-X\beta) \quad \text{.......21}$$

Differentiating (21) with respect to $\boldsymbol{\beta}$ will be denoted as $\mathbf{l}_{\beta}$ so we get

$$\boldsymbol{\ell}_{\beta} = \frac{\partial \ell}{\partial \beta} = \mathbf{X'V^{-1}}y - \mathbf{X'V^{-1}}X\boldsymbol{\beta} \quad \text{.......22}$$

Now differentiating equation (21) with respect to σ_i^2 and making use of:

$\frac{\partial V}{\partial \sigma_i^2} = \mathbf{Z_i Z_i'}$ and $\frac{\partial}{\partial \sigma_i^2}\log|V| = \mathbf{V^{-1}}\frac{\partial V}{\partial \sigma_i^2}\mathbf{V}^{-1}$ for i = 0, 1, 2, ... , r as a notational convenience is to define $u_o = \varepsilon$, $q_o = N$, $Z_o = I_N$ and $\sigma_\varepsilon^2 = \sigma_o^2$, leads to

$$\ell\sigma_i^2 = \frac{\partial l}{\partial \sigma_i^2} = \frac{-1}{2}\mathbf{tr}\ (\mathbf{V}^{-1}Z_iZ_i') + \frac{1}{2}(\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta})'\mathbf{V}^{-1}Z_iZ_i'\mathbf{V}^{-1}\ (\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta}) \quad \text{.......... 23}$$

Searle et al (2006, p. 236) then gives the ML equations as

$$\mathbf{X}'\dot{\mathbf{V}}^{-1}X\dot{\boldsymbol{\beta}} = \mathbf{X}'\dot{\mathbf{V}}^{-1}y \quad \text{.......... 24}$$

and

$$\mathbf{tr}\ (\mathbf{V}^{-1}Z_iZ_i') = (\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta})'\mathbf{V}^{-1}Z_iZ_i'\mathbf{V}^{-1}\ (\mathbf{y} - \mathbf{X}\,\boldsymbol{\beta}) \ \text{ for } i = 0, 1, 2, \ldots, r \quad \text{.......... 25}$$

An algebraically simpler expression for (24) is derived by defining

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\,\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} \quad \text{.......... 26}$$

Then from (23) it is clear that for $\dot{\boldsymbol{P}}$ being **P** with replaced by $\dot{\boldsymbol{V}}$

$$\dot{\mathbf{V}}^{-1}(y - X\dot{\boldsymbol{\beta}}) = \dot{\boldsymbol{P}}y \quad \text{.......... 27}$$

So that the ML equations (23) and (24) are

$$\mathbf{X}'\dot{\mathbf{V}}^{-1}X\dot{\boldsymbol{\beta}} = \mathbf{X}'\dot{\mathbf{V}}^{-1}y \quad \text{.......... 28}$$

and

$$\mathbf{tr}\ (\dot{\mathbf{V}}^{-1}Z_iZ_i') = \mathbf{y}'\ \dot{\boldsymbol{P}}Z_iZ_i'\ \dot{\boldsymbol{P}}\mathbf{y} \ \text{ for } i = 0, 1, 2, \ldots, r \quad \text{.......... 29}$$

where tr $(\dot{\mathbf{V}}^{-1}Z_iZ_i')$ is a column vector and $\dot{\mathbf{V}}$ is estimate of **V** where $\dot{\sigma}^2$ is used, in place of σ^2.

3.4.1.3.2. The Maximum Likelihood Solutions Using BLUP

Searle et al (2006, pp. 278-279) give the following solutions to the ML equations, that with the superscript (m) denoting computed values after *m* iterations. We have:

$$\boldsymbol{\sigma}_{\varepsilon}^{2(\mathbf{m+1})} = \frac{[y'(y - X\beta^{o(m)} - Z\tilde{u}^{(m)})]}{\mathrm{N}} \quad \text{.......... (30a)}$$

$$\sigma_{\mathrm{i}}^{2(\mathrm{m+1})}=\frac{\tilde{\mathrm{u}}'^{(\mathrm{m})}\tilde{\mathrm{u}}^{(\mathrm{m})}+\sigma_{\mathrm{i}}^{2(\mathrm{m})}\,\mathrm{tr}(\mathrm{W_{ii}}^{(\mathrm{m})})}{q_i}=\frac{\tilde{\mathrm{u}}'^{(\mathrm{m})}\tilde{\mathrm{u}}^{(\mathrm{m})}}{q_{i-\mathrm{tr}(\mathrm{W_{ii}})^{(\mathrm{m})}}} \quad \text{.......................................} (30b)$$

where $\mathbf{W}=(\mathbf{I}+\mathbf{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z}\boldsymbol{G})^{-1}=\mathrm{W_{ij}}$, i,j = 1, 2, 3, ..., r and G has q_i diagonal elements of σ_i^2 .

Searle et al (2006, pp.284-285) give the working of ML estimation via iterating as follows:

Consider the set of equations as above:

$$\sigma_{\varepsilon}^{2(\mathrm{m+1})}=\frac{[y'(y-X\beta^{(m)}-Zu^{(m)})]}{\mathrm{N}} \quad \text{..} (30a')$$

$$\sigma_{\mathrm{i}}^{2(\mathrm{m+1})}=\frac{\tilde{\mathrm{u}}'^{(\mathrm{m})}\tilde{\mathrm{u}}^{(\mathrm{m})}+\sigma_{\mathrm{i}}^{2(\mathrm{m})}\,\mathrm{tr}(\mathrm{W_{ii}})^{(\mathrm{m})}}{q_i}=\frac{\tilde{\mathrm{u}}'^{(\mathrm{m})}\tilde{\mathrm{u}}^{(\mathrm{m})}}{q_{i-\mathrm{tr}(\mathrm{W_{ii}})^{(\mathrm{m})}})} \quad \text{..} (30b')$$

1. Decide on starting values $\sigma_{\varepsilon}^{2(0)}$ and $\sigma_{\mathrm{i}}^{2(0)}$, and set m=0.
2. Calculate $W^{(m)}=(\sigma_{\varepsilon}^{2(\mathrm{m})}I+Z'ZG^{(m)})^{-1}\sigma_{\varepsilon}^{2(\mathrm{m})}$, (remember var (u) = G) and now solve for $\beta^{(m)}$ and $V^{(m)}$ and then calculate

$$\boldsymbol{u}^{(m)}=\boldsymbol{G}^{(m)}\boldsymbol{V}^{(m)} \text{ from } \begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{G}^{(m)} \\ \boldsymbol{Z}'\boldsymbol{X} & \boldsymbol{W}^{(m)} \end{bmatrix}\begin{bmatrix} \boldsymbol{\beta}^{(m)} \\ \boldsymbol{V}^{(m)} \end{bmatrix}=\begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{y} \end{bmatrix}$$

3. Now calculate $\sigma_{\varepsilon}^{2(\mathrm{m+1})}$ and either of the expressions for $\sigma_{\mathrm{i}}^{2(\mathrm{m+1})}$.
4. If convergence is reached for σ^2 's, set $\sigma^{2(m+1)}$. Denote the resulting terms as $\widetilde{W}$ $=W^{(m+1)}$, $\tilde{\beta}=\beta^{(m+1)}$, $\tilde{V}=V^{(m+1)}$ $and\ \tilde{u}=\mathrm{u}^{(m+1)}$. Use $\tilde{\sigma}^2$ and $\widetilde{W}$ to calculate the information matrix, $\mathrm{I}(\tilde{\sigma}^2_{\mathrm{ML}})$.
5. If convergence is not reached, increase m by unity and return to step 2. At each repeat of step 3 uses whichever of the equations (a) or (b), which was used on the first occasion.

The matrix V is by definition always non-negative definite and usually positive definite. Therefore, the ML estimators σ_{ε}^2 and σ_j^2 must satisfy these constraints that $\sigma_{\varepsilon}^2>0$ and

$\sigma_j^2 \geq 0$ for i = 1, 2, 3, ..., r. Computer programs must be able to satisfy their constraints and we replace the negative solution by a small positive number if any σ_j^2 be set to zero.

The dispersion matrix of the parameter estimates is estimated using the inverse of the information matrix, by substituting the maximum likelihood equations for the parameters in the information matrix. Longford (1993) points out that this becomes problematic for small data sets since; (a) the asymptotic properties may not apply them and (b) there is a lot of uncertainty about the parameters involved in the information function and the information function may vary substantially with the parameters.

Finally to lead us on to the next section in this chapter, one of the criticisms leveled at ML, is the statistical analyses used so far in cereal research as related to various factors were ordinary maximum likelihood estimator of variance component. But this takes NO account of the degree of freedom used in estimating fixed effect and hence is biased (as cited by Girma, 2005).

3.4.1.4. Restricted maximum likelihood (REML)

A major drawback of ML estimation in a mixed model is that fixed effects are treated as if they were known, i.e. the loss in degrees of freedom due to fitting these effects is ignored. Fortunately, a modified ML procedure, the so-called Restricted (Marginal) Maximum Likelihood (REML) as described by Patterson and Thompson (1971), overcomes this problem by maximizing only part of the likelihood which is independent of the fixed effects. Conceptually, this is achieved by replacing the data by linear functions thereof, **'error contrasts'**, with an expectation of zero. These can be viewed as the observations adjusted for generalized least-squares estimates of the fixed effects. One of the attractive features of REML is that it takes into account the degree of freedom in the variance component estimation and it is based on the idea of estimating the variance components via the residuals calculated after fitting by ordinary least squares of just the fixed effects part of the model.

Even more than ML, REML estimation of variance components is computationally highly demanding and this has limited practical applications. However, over the last decade considerable research effort has concentrated on the development of specialized and efficient algorithms. Advances in theory, in particular the development of specialized and efficient algorithms, together with an increase in the general level of computing power available have led to progressive use of REML. Widely distributed statistical packages like SAS now provide options for REML analyses.

3.4.1.4.1. The Basis

REML is often interpreted as a technique that is based on linear combinations of y, not forgetting that these linear combinations do not contain any fixed effects. Not surprisingly these linear combinations of values not containing any fixed effects turn out to be equivalent to the residuals obtained after we fit the model the fixed effects. Consider the set of values **C'y** where matrices of the form **C'** can be chosen to satisfy

$\mathbf{C'y = C'X}\beta + \mathbf{C'Zu}$ such that no term in $\boldsymbol{\beta}$ is contained. i. e.;

$$\mathbf{C'X}\beta = \mathbf{0} \;\; \forall \beta \quad \text{..........} 31$$

$$\Rightarrow \mathbf{C'X = 0}, \text{ with } \mathbf{C'}_{rxN} \text{ of rank } r = r(X)$$

$$\mathbf{P} = V^{-1} - V^{-1}\mathbf{X}(\mathbf{X'}V^{-1}\mathbf{X})^{-}\mathbf{X'}V^{-1} = \mathbf{C}\,(\mathbf{C'}V^{-1}\mathbf{C})^{-1}\mathbf{C'} \quad \text{..........} 32$$

3.4.1.4.2. REML Equations

Searle et al (2006, pp. 251- 252) give the REML equations which we summarize as follows: We have that $\mathbf{y} \backsim \mathbf{N}\,(\mathbf{X}\beta, V)$ for $\mathbf{C'X = 0}$ so $\mathbf{C'y} \backsim \mathbf{N(0, C'VC)}$.

Using the Maximum Likelihood estimation equation we have that

$$\mathrm{tr}(\dot{\mathrm{V}}^{-1}\mathrm{Z_i Z_i'}) = \mathrm{y'}\dot{\mathrm{P}}Z_i\mathrm{Z_i'}\dot{\mathrm{P}}\mathrm{y}, \; i = 1, 2, 3, ..., r. \quad \text{..........} 33$$

So if we make the following replacements: Replace **y** by **C'y, X** by **C'X = 0, Z** by **C'Z** and **V** by **C'VC**. Then equation (30) becomes:

$$\mathbf{tr}\ [(C'\dot{V}C)^{-1}C'Z_iZ_i'C] = y'C(C'\dot{V}C)^{-1}C'Z_iZ_i'C\ (C'\dot{V}C)^{-1}C'y,\ i= 1, 2, 3,\ldots, r. \quad \ldots\ldots 34$$

And $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^{-}\mathbf{V}^{-1} = \mathbf{C}(\mathbf{C'VC})^{-1}\mathbf{C'}$ so that equation (34) now becomes

$$\mathbf{tr}\ (\dot{P}Z_iZ_i') = y'\dot{P}\ Z_iZ_i'\dot{P}y,\ i= 1, 2, 3, \ldots\ldots r. \quad \ldots\ldots 35$$

It is clear that **PVP = P** so we can use the following identity to give us an alternate form of the REML equations (as given by Searle et al 2006)

$$\mathbf{tr}\ (Z_i'\dot{P}Z_jZ_j'\dot{P}Z_i)\dot{\sigma}^2 = y'\dot{P}\ Z_iZ_i'\dot{P}y,\ i= 1, 2, 3,\ldots, r. \quad \ldots\ldots 36$$

Searle et al (2006, p.252) state that the REML equations don't contain K except only through its relationship to **P.**

Thus we can express **P** as $\mathbf{P} = V^{-1} - V^{-1}\mathbf{X}(\mathbf{X'}V^{-1}\mathbf{X})^{-}\mathbf{X'}V^{-1} = \mathbf{C}\ (\mathbf{C'}V^{-1}\mathbf{C})^{-1}\mathbf{C'}$ which does not involve **C**. This shows that the REML equations are invariant to a particular set of error contrasts that are chosen.

Searle et al (2006, pp. 282-284) gives the following solutions to the REML equations, that with the superscript (m) denoting computed values after *m* rounds of iteration. We have:

$$\sigma_{\varepsilon}^{2(m+1)} = \frac{[y'(y - X\beta^{o(m)} - Z\tilde{u}^{(m)})]}{N-r} \quad \ldots\ldots (37)$$

$$\sigma_i^{2(m+1)} = \frac{\tilde{u}'^{(m)}\tilde{u}^{(m)} + \sigma_i^{2(m)}\ \mathrm{tr}(T_{ii}{}^{(m)})}{q_i} = \frac{\tilde{u}'^{(m)}\tilde{u}^{(m)}}{q_{i-\mathrm{tr}(T_{ii})^{(m)}}} \quad \ldots\ldots (38)$$

Where $\mathbf{T} = (\mathbf{I} + \mathbf{Z'SZD})^{-1} = \mathbf{T}_{ij}$, i,j = 1, 2, 3, ..., r and D has q_i diagonal elements of σ_i^2.

3.5. Estimation of the Parameters

After a convergence criterion is fulfilled, then the next point is estimation of the parameters. For the mixed model given by equation (2), a key assumption in the foregoing analysis is that U and $\boldsymbol{\varepsilon}$ is normally distributed with $\mathrm{E}\begin{bmatrix}\mathbf{U}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}$ and $\mathbf{var}\begin{bmatrix}\mathbf{U}\\ \boldsymbol{\varepsilon}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}$.

Hence, the variance of y is, therefore, **V = ZGZ' + R**. As a result we can model V by setting up the random effects design matrix **Z** and by specifying covariance of **G** and **R**, with **Z** containing dummy variables, **G** containing variance components in a diagonal structure, and $\mathbf{R} = \sigma_\varepsilon^2\mathbf{I}_N$, where $\mathbf{I}_N$ denotes the $N \times N$ identity matrix. As it is shown above, for the foregoing analysis we need to know **G** and **R**, since most of the time they are unknown parameters, we first find estimates of those parameters.

3.5.1. Estimating G and R in the Mixed Model

Estimation of parameters in the mixed model is more difficult than in the general linear model. Because there are not only needs fixed effect parameter estimates as in the general linear model, but also we have unknown parameters in **u**, **G**, and **R** as well. Generalized least squares (GLS) is more appropriate than Least Square (LS) and applied by minimizing $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'V^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ or maximize the multivariate normal distribution function after taking the natural logarithm. However, it requires knowledge of **V** and, therefore, knowledge of **G** and **R**. Lacking such information, one approach is to use estimates *GLS*, in which we insert some reasonable estimate for **V** into the minimization problem. The goal thus becomes finding a reasonable estimate of **G** and **R** and intern **V**. In many situations, the best approach is to use likelihood-based methods, exploiting the assumption that **u** and ϵ are normally distributed (Hartley and Rao, 1972; Patterson and Thompson 1971; Harville 1977; Laird and Ware 1982; Jennrich and Schluchter 1986). PROC MIXED in SAS implements two likelihood-based methods: maximum likelihood (ML) and restricted/residual maximum likelihood (REML). Using calculus, it is possible to reduce this maximization problem to one over only the parameters in **G** and **R**. The corresponding log-likelihood functions are as follows:

$$\begin{aligned}
\text{ML:} \quad l(\mathbf{G},\mathbf{R}) &= -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\mathbf{r}'\mathbf{V}^{-1}\mathbf{r} - \frac{n}{2}\log(2\pi) \\
\text{REML:} \quad l_R(\mathbf{G},\mathbf{R}) &= -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| \\
&\quad -\frac{1}{2}\mathbf{r}'\mathbf{V}^{-1}\mathbf{r} - \frac{n-p}{2}\log(2\pi)\}
\end{aligned}$$

where $\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ and p is the rank of **X**.

Mixed model actually minimizes -2 times these functions using a ridge-stabilized Newton-Raphson algorithm. One advantage of using the Newton-Raphson algorithm is that the second derivative matrix of the objective function evaluated at the optima is available upon completion. Denoting this matrix **H**, the asymptotic theory of maximum likelihood (Serfling 1980) shows that $2\mathbf{H}^{-1}$ is an asymptotic variance-covariance matrix of the estimated parameters of **G** and **R**. However, these can be unreliable in small samples, especially for parameters such as variance components which have sampling distributions that tend to be skewed to the right. But since, a residual variance σ^2 is a part of mixed model; it can usually be profiled out of the likelihood. This means solving analytically for the optimal σ^2 and plugging this expression back into the likelihood formula (Wolfinger et al, 1994). This reduces the number of optimization parameters by one and can improve convergence properties. Mixed model profiles the residual variance out of the log likelihood whenever it appears reasonable to do so. Therefore, in Mixed Model analysis, the ML,REML, or Type I through Type III provide estimates of **G** and **R**, which are denoted $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{R}}$,respectively.

3.5.2. Estimating Fixed Effects and Predicting Random Effects in Mixed Model

Inferences about fixed effects have come to be called estimates, whereas those that concern random effects are known as predictions. Procedures for obtaining such estimators and

predictors have been developed using a variety of approaches. The most widely used procedures are BLUE and BLUP, referring respectively to best linear unbiased estimator and best linear unbiased predictor. They are best in the sense that they minimize the sampling variance, linear in the sense that they are linear functions of the response variable, and unbiased in the sense that E[BLUE($\hat{\beta}$)] = β and E[BLUP($\hat{U}$)] = E(U).
For the mixed model given by Equation (2), the BLUE of β is:

$$\hat{\beta} = (X'V^{-1}X)^{-1}\,X'V^{-1}y \qquad \text{.......39}$$

with V = ZGZ' +R provided all indicated inverses exist. If not, then generalized inverses are used and this is just the generalized least squares (GLS) estimator. Henderson (1963) showed that the BLUP of **U** is:

$$\hat{U} = GZ'V^{-1}(y - X\hat{\beta}) \qquad \text{.......40}$$

which is equivalent to the conditional expectation of **u** given **y** under the assumption of multivariate normality and since everything is Gaussian, these are linear functions of the data, and as everything is linear, they are unbiased. They have minimum variance amongst such estimators. The solution of Equations (37) and (38) requires the inverse of the covariance matrix **V.** However, the computation of V^{-1} can be quite difficult. As a way around this problem, Henderson (1984) offered a more compact method for jointly obtaining $\hat{\beta}$ and $\hat{U}$ in the form of his mixed-model equations (MME),

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ X'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \qquad \text{.......41}$$

While these expressions may look considerably more complicated than Equations (37) and (38), R^{-1} and G^{-1} are trivial to obtain if **R** and **G** are diagonal, and hence the sub-matrices in Equation (39) are much easier to compute than V^{-1}. A second advantage of Equation (39) can be seen by considering the dimensionality of the matrix on the left; the matrix that needs to be inverted to obtain the solution for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{U}}$ is of order $(p+q)$ x $(p+q)$, which is usually considerably less than the dimensional of **V** (an *NxN* matrix).

Although there are several ways to derive the mixed-model equations (Robinson 1991), Henderson (1953) originally obtained them by assuming that the covariance matrices **G** and **R** are known and that the densities of the vectors **u** and $\boldsymbol{\varepsilon}$ are each multivariate normal with no correlations between them. Equation (39) then yields the maximum likelihood estimates of the fixed and random effects. Henderson (1963) later showed that the mixed-model equations do not actually depend on normality, and that $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{U}}$ are BLUE and BLUP, respectively, under general conditions provided the variances are known. However, the contrary is usually true, so that (37) and (38) cannot be used in their basic forms. In most cases, $\mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$ is a function of a vector of covariance parameters, $\boldsymbol{\theta}$. The estimated least squares approach first parameter estimates of the covariance matrices **G** and **R**, then substitutes these estimates for the parameters into the functional form of $\mathbf{G}(\theta)$ and $\mathbf{R}(\theta)$ to obtain $\widehat{G} = \mathbf{G}(\theta)$ and $\widehat{R} = \mathbf{R}(\theta)$. Then $\widehat{\mathbf{V}}$ is computed as $\widehat{V} = \mathbf{Z}\widehat{G}\ \mathbf{Z}' + \widehat{R}$. Maximum likelihood, Residual maximum likelihood, or Type I, Type II, Type III sum of Squares has become the favored method of estimation (Searle et al 2006) of random effect parameters (**G** and **R**).

Thus, an estimate $\widehat{V}$ of **V** must be used in the computation, providing the "estimated" generalized least squares (EGLS) estimate (BLUE) of $\boldsymbol{\beta}$ and BLUP are becomes:

$$\widetilde{\beta} = (X'\widehat{V}^{-1}X)^{-1} X'\widehat{V}^{-1}y \quad \text{.......(37')}$$

and

$$\widetilde{U} = \widehat{G}Z'\widehat{V}^{-1}(y - X\widetilde{\beta}) \quad \text{.......(38')}$$

3.5.2.1. Standard Errors

Relatively straightforward extension of Henderson's mixed-model equations provides estimates of the standard errors of the fixed and random effects. Let the inverse of the left most matrices in Equation (39) are:

$$\widehat{\mathbf{C}} = \begin{bmatrix} \mathbf{X}'\widehat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\widehat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{Z} + \widehat{\mathbf{G}}^{-1} \end{bmatrix}^{-}$$

.... 42

McLean and Sanders (1988) show that $\widehat{\mathbf{C}}$ can also be written as

$$\widehat{\mathbf{C}} = \begin{bmatrix} \widehat{\mathbf{C}}_{11} & \widehat{\mathbf{C}}'_{21} \\ \widehat{\mathbf{C}}_{21} & \widehat{\mathbf{C}}_{22} \end{bmatrix}$$

....43

where,

$$\begin{aligned} \widehat{\mathbf{C}}_{11} &= (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-} \\ \widehat{\mathbf{C}}_{21} &= -\widehat{\mathbf{G}}\mathbf{Z}'\widehat{\mathbf{V}}^{-1}\mathbf{X}\widehat{\mathbf{C}}_{11} \\ \widehat{\mathbf{C}}_{22} &= (\mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{Z} + \widehat{\mathbf{G}}^{-1})^{-1} - \widehat{\mathbf{C}}_{21}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{Z}\widehat{\mathbf{G}} \end{aligned}$$

....44

and $\widehat{C}_{11}$, $\widehat{C}_{21}$, and $\widehat{C}_{22}$ respectively, are pxp, qxp, and qxq sub matrices. Using this notation, Henderson (1975) showed that the sampling covariance matrix for the BLUE of $\beta$is given by $\sigma(\widehat{\beta})$ = $\widehat{C}_{11}$, that the sampling covariance matrix of the prediction errors $(\widehat{u} - u)$ is given by σ $(\widehat{\boldsymbol{u}} - \boldsymbol{u})$ = $\widehat{C}_{22}$, and that the sampling covariance of estimated effects and prediction errors is given by σ $(\widehat{\beta}, \widehat{u} - u)$ = $\widehat{C}_{12}$. The standard errors of the fixed and random effects are obtained, respectively, as the square roots of the diagonal elements of $\widehat{C}_{11}$ and $\widehat{C}_{22}$.

For inferences concerning the fixed- and random-effect parameters in mixed model, consider estimable linear combinations of the form L$[{\beta \atop U}]$ (SAS 2000). The estimability requirement (Searle et al 2006) applies only to the β-portion of L, as any linear combination of U is estimable.

The GLS estimate provides best unbiased estimates $L\widehat{\beta}$ of estimable linear combinations $L\beta$ of the elements of the fixed effect parameter vector.

The covariance matrix of $\widehat{\boldsymbol{\beta}}$ is $\mathbf{V}(\widehat{\boldsymbol{\beta}}) = (\mathbf{X}'\widehat{V}^{-1}X)^{-1}$ and the covariance matrix of a set of linear combinations $\mathbf{L}\widehat{\boldsymbol{\beta}}$ is $\mathbf{V}(L\widehat{\boldsymbol{\beta}}) = \mathbf{L}(\mathbf{X}'\widehat{V}^{-1}X)^{-1}L'$, which provides the machinery for statistical inference regarding $L\beta$ specifically, a test of null hypothesis $\mathbf{H_0} : \mathbf{L}\boldsymbol{\beta} = \mathbf{0}$ is given by the statistic $\widehat{\boldsymbol{\beta}}'L'\ [\mathbf{L}(\mathbf{X}'\widehat{V}^{-1}X)^{-1}\mathbf{L}]^{-1}\mathbf{L}\widehat{\boldsymbol{\beta}}$, which has a chi-square distribution with degrees of freedom r(L), the rank of L. A 100(1 - α) % confidence interval for a single linear combination $\mathbf{l}\boldsymbol{\beta}$ is given by $\mathbf{1}\widehat{\boldsymbol{\beta}} = z_{\alpha/2}\mathbf{1}(\mathbf{X}'\widehat{V}^{-1}X)^{-1}\ \mathbf{1}$, where $z_{\alpha/2}$ is the 1 - $\alpha/2$ percentile of the standard normal distribution.

3.6. Model Diagnostics

A statistical model, whether of the fixed-effects or mixed-effects factors, represents how we think the data were generated. Diagnostics is the assessment of agreement of the model and the data fitted and it is very important in linear mixed models because likelihood based estimation methods are particularly sensitive to unusual observations especially. Many graphical methods and analytical techniques for linear regression model extend to mixed model setting as well.

3.6.1. Influence Diagnostics in the Mixed Model

It is well known that the maximum likelihood and /or restricted maximum likelihood estimates of the variance ratios (or components), fixed and random effects, and the likelihood functions can be substantially influenced by one or a few observation(s); that is, not all observations are of equal importance. It is therefore important for the analyst to be able to identify such observations and assess their effects on various aspects of the analysis. To this end, several diagnostics have been proposed (Beckman et al., 1987; Christensen et al., 1992; Hocking et al., 1989; Lesaffre and Verbeke, 1998). Recently, Zewotir and Galpin (2005) extended the work of Christensen et al (1992) and found computationally efficient formulae, which are similar to influence measures in standard linear regression.

3.6.1.1. Overall Influence

An overall influence measure is the likelihood distance (Cook and Weisberg 1982, Ch. 5.2). The basic procedure for quantifying influence is simple: (a) Fit the model to the data and obtain estimate of all parameters, (b) remove one or more data points from the analysis and compute updated estimates of model parameters, (c) based on full- and reduced-data estimates, contrast quantities of interest to determine how the absence of the observations changes the analysis.

Then the likelihood and restricted likelihood distances are obtained as:

$$LD_i = 2[l_{full\ data\ estimate} - l_{reduced\ data\ estimate}],\ i = 1, 2, 3, \ldots, N \quad \text{.......45}$$

The likelihood distance gives the amount by which the log-likelihood of the full data changes if one were to evaluate it at the reduced-data estimates and a large value of LD_i indicates that observation(s) are influential. The likelihood distance is a global, summary measure, expressing the influence of the observations in the set of reduced data on all parameters in that were subject to updating. If the global measure suggests that the points in reduced data are influential, we should next determine the nature of that influence. In particular, the points can affect (i) the estimates of fixed effects (ii) the estimates of the precision of the fixed effects (iii) the estimates of the covariance parameters (iv) the estimates of the precision of the covariance parameters and (v) fitted and predicted values.

3.6.1.2. Change in Parameter Estimates

Because the number of fixed-effects and covariance parameters can be large, the mixed procedure enables experimenters to compute summary statistics that capture the change in the entire parameter vector. These quadratic forms are based on Cook's Distance (Cook

1977) and the multivariate DFFITS statistic, (MDFFITS; Belsley, Kuh, and Welsch 1980, p. 32). For both statistics, we are concerned about large values, indicating that the change in the parameter estimate is large relative to the variability of the estimate. For the fixed effect the two statistics are:

$$\mathbf{D}(\boldsymbol{\beta})_i = \frac{(\widehat{\boldsymbol{\beta}}_{full\ data} - \widehat{\boldsymbol{\beta}}_{reduced\ data})' \operatorname{var}(\widehat{\boldsymbol{\beta}}_{full\ data})^{-1} (\widehat{\boldsymbol{\beta}}_{full\ data} - \widehat{\boldsymbol{\beta}}_{reduced\ data})}{rank(X)}$$

$$\mathbf{DFFIT}_i = \frac{(\widehat{\boldsymbol{\beta}}_{full\ data} - \widehat{\boldsymbol{\beta}}_{reduced\ data})' \operatorname{var}(\widehat{\boldsymbol{\beta}}_{reduced\ data})^{-1} (\widehat{\boldsymbol{\beta}}_{full\ data} - \widehat{\boldsymbol{\beta}}_{reduced\ data}}{rank(X)} \quad \text{............46}$$

If the covariance parameters are updated during influence analysis, similar statistics can be computed for covariance estimates.

3.6.1.3. Change in Precision of Estimates

The effect on the precision of estimates is separate from the effect on the point estimates. Data points that have a small Cook's Distance, for example, can still greatly affect hypothesis tests and confidence intervals, if their influence on the precision of the estimates is large. Zewotir and Galpin (2005) compute functions of the trace and determinants of the variance matrices based on the full-data and the reduced-data estimates:

$$\text{COVTRACE }(\beta) = \left| \text{trace }\{(\widehat{var}[\hat{\beta}_{full\ data}])^{-1} \widehat{var}[\hat{\beta}_{reduced\ data}]\} - \text{rank(X)} \right|$$

$$\text{COVRATIO }(\beta) = \frac{det(\widehat{var}[\widehat{\boldsymbol{\beta}}_{reduced\ data}])}{\det(\widehat{\text{var}}[\widehat{\boldsymbol{\beta}}_{full\ data}])} \quad \text{............47}$$

The benchmarks of "no influence" are zero for the covariance trace and one for the covariance ratio. If the influence analysis updates the covariance parameters, the mixed procedure computes similar statistics for covariance estimates. That is:

COVTRACE (θ) = $\left| \text{trace}\ \{(\widehat{var}[\hat{\theta}_{full\ data}])^{-1}\widehat{var}[\hat{\theta}_{reduced\ data}]\} - q \right|$

COVRATIO (θ) = $\dfrac{det(\widehat{var}[\hat{\theta}_{reduced\ data}])}{\det(\widehat{\text{var}}[\hat{\theta}_{full\ data}])}$ 48

where, q denotes the rank of variance components. The computation of COVTRACE and COVRATIO for covariance parameters is obtained from the inverse Hessian matrix using PROC MIXED of ASYCOV. After filtering the data that influences the model promote to choosing covariance structure and model selection in the next section.

3.6.2. Choosing Covariance Structure and Model selection

We have many choices for the covariance structures (**G** and **R**). Ideally the covariance structure should be known from previous work or subject matter condition. Otherwise, one should risk of shopping for the structure that leads to a better fit. We contemplate a few life structures and choose among them according to some measures of fit. These lead to be composed as two measures; one that rewards for the acquiring of the fit and another penalize for the number of parameters it takes to achieve best fit.

- Reward: looks at how well the estimated and observed structures agree.
- Penalty: considers how many parameters it takes to achieve the fit.

Information criteria such as AIC (Akaike, 1974), BIC (Schwarz, 1978), AICC (Hurvich and Tsai, 1989), and CAIC (Bozdogan, 1987) are available for selecting the covariance structure (***G***, and ***R***) and also for model selection. In general, these information criteria are functions of both the maximized log-likelihood for a given model (l) and a penalty term based on the number of parameters (q) in the model (Gurka, 2006). Common formulas for AIC and BIC are:

$$\text{AIC} = -2\ell + 2q$$

and $$\text{BIC} = -2\ell + q\,(\log N^{*})$$ 49

Where, ℓ is the log-likelihood of either of ML or REML, N^{*} = N under ML, and N^{*} = N − p under REML. For ML estimation, q = p + k, the total number of parameters in the model.

However, under REML estimation, q = k, the number of covariance parameters, and p is the number of fixed effects parameters (Gurka, 2006). Note that the AICC corrects for small sample sizes and as the sample size increases, the AICC converges to the AIC. In the same way, as the sample size increases, the CAIC converges to the BIC (SPSS results coach, Version 13.0). The model with the smaller value of either of the above is the better the model fits for the data. But, since we are interested in getting as simple models as possible we also have to consider the number of parameters in the structures which is the smallest as much as possible.

3.6.3. Model Checking

As we have appropriate covariance structure for the selected model, it is trivial to check whether or not the error term for the linear mixed model is assumed to be independent and normally distributed with zero mean and constant variance. In other words check whether or not the observed residual fitted would have mean zero and constant variance. There are many mechanisms to check the normality assumptions of the model, like the normal P-P plots, histogram of residuals, and a scatter plot of standardized residuals versus predicted values. All these are strengthen using SAS PROC UNIVARIATE procedure for residuals.

Normal probability plot (P_P plot) is sketched by using each residual against the expected value under normality. A plot that is nearly linear suggests that the normality assumption is valid, whereas a plot departs substantially from linearity suggestion that the distribution of the errors is not normal. If the sketch of normal P-P plot is scattered around the straight line, then we can say the observed error satisfies the stated normality assumptions. Substantial departure from the straight line indicates that the distribution is not normal. On the other hand, if the plot shows a certain pattern instead of following the straight line randomly then we can conclude that the samples are taken from either positive or negative skewed distribution based on the pattern of plot (Douglas et al 1991).

The normality assumption may also be checked by constructing histogram of residuals. However, if the number of residuals is too small it is too difficult to allow easy visual identification of the shape of the normal distribution. If the line of normal curve is almost symmetric around the mean of the residual, then there is an indication of the satisfaction of the normality assumption. However, if the normal curve is tailed in either left or right, the assumption of normality is failed.

The standardized residual[3] are also useful tool in detecting departures of the error term from the normality. A plot of standardized residual against the corresponding fitted (predicted) values of the dependent variable is useful to such checking. If the plot of standardized residual versus predicted values lie within plus or minus two horizontal bands, then there are no model violations of normality assumptions.

In general, if the errors are normally distributed, then approximately 99 percent of them should fall between plus and minus three. If the scatter plot of standardized residual versus predicted values lie outside the specified horizontal band with the large number of observation, then it is possible to say that there is a model deficiency (Douglas et al 1991).

3: The standardized residual is calculated either $d_i = \frac{\varepsilon_i}{\sqrt{MSE}}$, stand. or $r_i = \frac{\varepsilon_i}{\sqrt{MSE\left[1-\left(\frac{1}{n}+\frac{(X_{i-}\bar{X})}{S_{xx}}\right)^2\right]}}$,stud.

CHAPTER FOUR: MODEL DIAGNOSTICS

4.1. Choosing Covariance Structures and model selection

In order to estimate variance components, firstly we need to obtain covariance structures and have a model having an adequate representation of the data. We have many structures for the variance components (G and R). Ideally, the covariance structure should be known from previous work or subject matter condition. Otherwise, one should take the risk of shopping for variance structure that leads to a better fit. We contemplate a few life structures and choose among them according to some measures of fit (Lindsey 1993). Here in this thesis, the data are not longitudinal and the structure of G may have VC, CS or UN. A model fitted based on variance component structure (VC) has smallest AIC among all the above covariance structures. In addition to this, VC has a simple covariance structure as it has one parameter (σ^2) only. This implies that VC is appropriate for analysis and further interpretation. In fact, covariance structure in this thesis is variance component (VC).

The PROC MIXED procedure also produces further output for model selection. AIC is used to compare models having the same fixed effects but different random effects (Akaike 1974). The BIC compares two models and selects smaller BIC even though it penalizes models with a greater number of covariance parameters more than AIC does (Schwarz 1978). Therefore, AIC and BIC are extremely valuable in comparing two or more models to assess which is the best model (fit). In addition to these, the difference in log-likelihoods between two nested models is the most basic measure for model selection (SPSS results coach, Version 13.0). The difference in log likelihood, therefore, gives a likelihood ratio test in the usual way where it is the test statistic which has an asymptotic chi-squared distribution with degree of freedom equal to the differences between parameters of the model. This study utilizes this procedure to check for random effects to the model, i. e., variety and variety*environment.

Table 4.2 shows that all models are comparable as convergence criteria are met with positive Hessian matrix. If convergence is met and the estimated Hessian matrices are positive definite (i.e., the variances are positive and the absolute value of implied correlations do not exceed 1.0), then estimators have some desirable properties. The fixed effect estimates are unbiased (Kackar and Harville, 1984) and estimates of variance parameters (elements in random effects and residual) are asymptotically unbiased (Raudenbush and Bryk, 2002). The estimates of fixed effects and variance parameters also tend to be asymptotically efficient. Taking this into consideration, a model that contained random variables variety and variety*environment has smallest values of AIC and BIC for ML (AIC =6884.5 and BIC = 6953.6). This model also has the smallest values of AIC and BIC for REML (AIC = 6773.3 and BIC = 6779.7). Moreover, the difference between -2loglikelihood of a model that ignores both variety and variety*environment and a model with all fixed effects (env, rep(env) and block(rep)) and random effects variety and variety*environment are 163.2 and 162 for ML and REML respectively. Thus, we have 9.21 for both ML and REML with 2 degree of freedom at 1% significance level of chi-squares critical value. Hence, we conclude that effects of variety and variety*environment are quite significant and should be included as random effects in the model. Therefore, a model with all fixed effects (env, rep(env) and block(rep)), random factors variety and variety*environment is appropriate to the data set of this study.

Table 4.2: Selected Information Criterion for model selections.

Variables	Methods								-2 LogLike Diff ML (REML)
	ML				REML				
	Criteria*				Criteria*				
	parms	-2 LogLike	AIC	BIC	parms	-2 LogLike	AIC	BIC	
Fixed effs	30	6983.7	7043.7	7191.6	1	6929.3	6931.3	6936.2	-
Fixed eff s & var	31	6853.3	6915.3	6982.2	2	6796.4	6800.4	6804.7	130.4 (132.2)
Fixed effs & var*env	31	6898.7	6960.7	7092.1	2	6847.8	6851.8	6860.3	85 (81.5)
Fixed effs, var & var*env	**32**	**6820.5**	**6884.5**	**6953.6**	**3**	**6767.3**	**6773.3**	**6779.7**	**163.2 (162)**

**In all model fitting the criterion is met.*

Now the linear mixed model from chapter 3 i. e., **y =X*β* + ZU +*ε*** can be fitted to the data as:

Y_{1024x1}: a vector representing yield of barley in quintal per hectare.

$X_{1024x41}$: the design matrix of fixed effects. β_{41x1}: vector parameters of fixed effects.

$Z_{1024x576}$: represents the incidence matrix for random effects (variety and variety*env).

U_{576x1}: represents the random effect parameters of variety and variety*env.

4.2. Influence Diagnostics

Table C1 (in the Annex) presents the result of reduced-data parameter estimates for the fixed effects and the covariance parameters arranged based on the restricted likelihood distances (RLD) with top 50 influence observations. Clearly, observation 140 is the largest in terms of restricted likelihood distance (RLD) followed by observation 434 as we see from Table C1 and Figure 1. The influence of this largest observation on the covariance parameters and its precision is large (Cook's D =0.26279 and COVRATIO=0.9108). This observation also exerts influence on the estimates of the fixed effects and their precision as can be seen from Cook's D and the covariance ratio (Cook's D = 0.01784 and COVRATIO =0.8436) statistic.

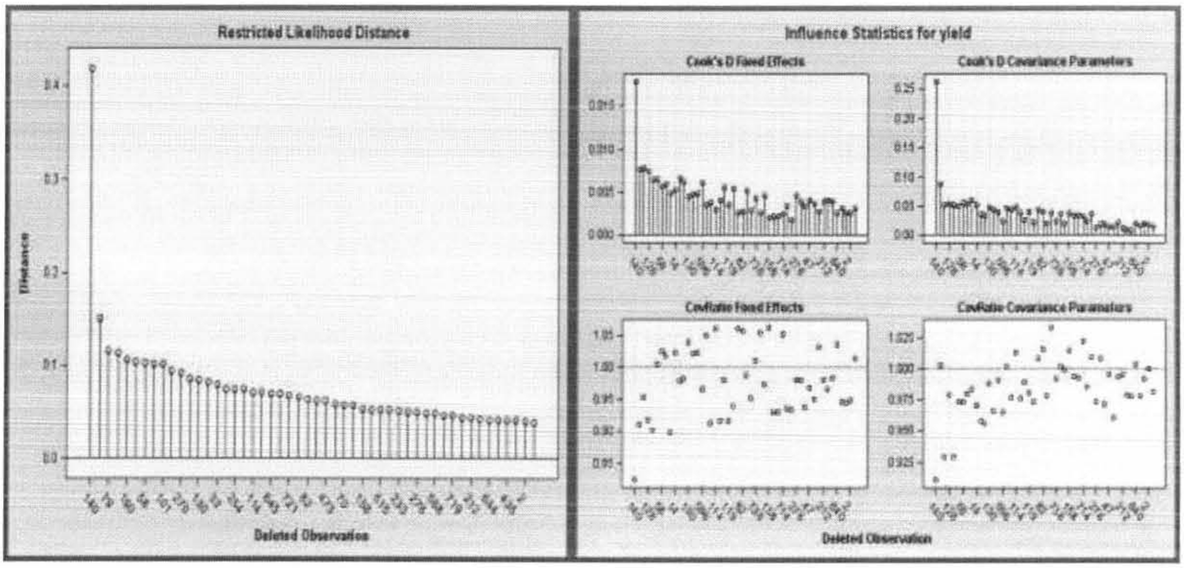

Figure1: Influential measures for fixed effects and covariance parameters by case deletion.

Zewotir and Galpin (2006) showed that the influence of the i[th] observation on the estimate of fixed effect parameters (or its variance) can be detected by Cook's distance, covariance-ratio and covariance-trace. As the estimate of fixed effects parameter is a function of response variable and fixed effects, outliers in response variable or in fixed effects are potentially influential points on the estimate of fixed effects parameter. Cook's distance is sensitive to the case of only an outlier in a response variable and this is masked if there is a perturbation in fixed effects in the same observation. Covariance-ratio and covariance-trace are sensitive to either high leverage points in fixed effects or a severe outlier in response variable, or both. In the same analogue, the influence of an observation on the prediction of random effects (or its variance) is likely to be detected by Cook's distance, covariance-ratio and covariance-trace of random effects. These are sensitive to an outlier in response variable. This sensitivity is enhanced and masked if there is a high leverage in random effects and fixed effects in the same observation, respectively. Therefore, we have seen masking effects from Figure 2 and it needs further works.

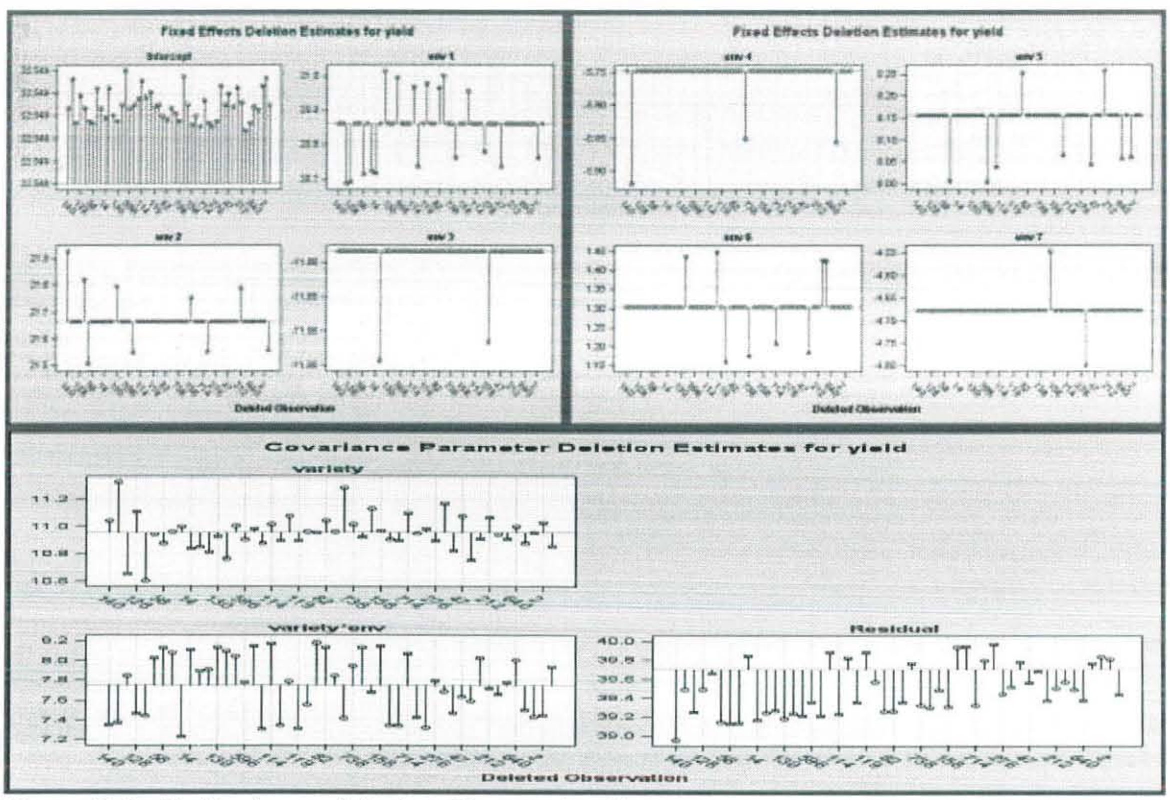

Figure 2: Individual case deletion diagnostics diagram.

In summary, there are masking effects in the diagnostic measures. This is inevitable whenever there are multiple outliers. Separate analysis for high-leverage points and outliers in the mixed model (Zewotir and Galpin, 2006) before influence diagnostics may minimize this masking effect. Multiple observations influence diagnostics is also important to detect observations which are jointly but not individually influential. Thus, multiple observations diagnostic in the linear mixed model is an area of future research.

4.3. Model Checking

As discussed in Section 6.3.2, we used the PP-plot of residual, histogram of residual, and the scatter plot of standardized residual versus predicted values to check whether the stated normality assumptions is met.

The PP-plot of residuals in Figure 3 shows that the data conforms to the hypothetical normality assumptions. The fact that the plot is scattered around the straight line and does not show considerable patter indicates that the distribution of the error term and the response variable is normal (linearity of the error term is fulfilled).

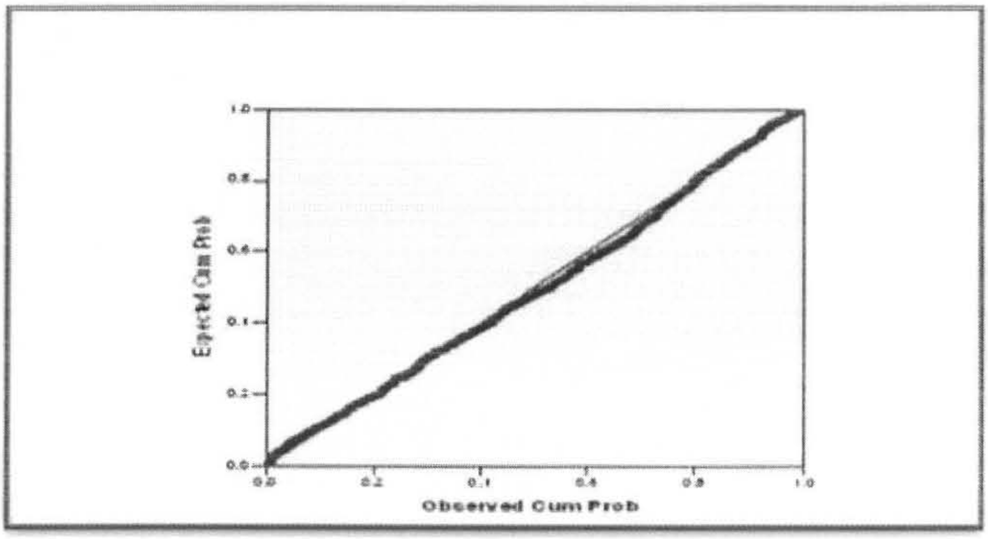

Figure3: The normal Probability plot of mixed effects for residuals.

To strengthen this conclusion, the histogram of residual is better than PP-plot to identify easily the shape of the normal distribution as the number of data is too large.

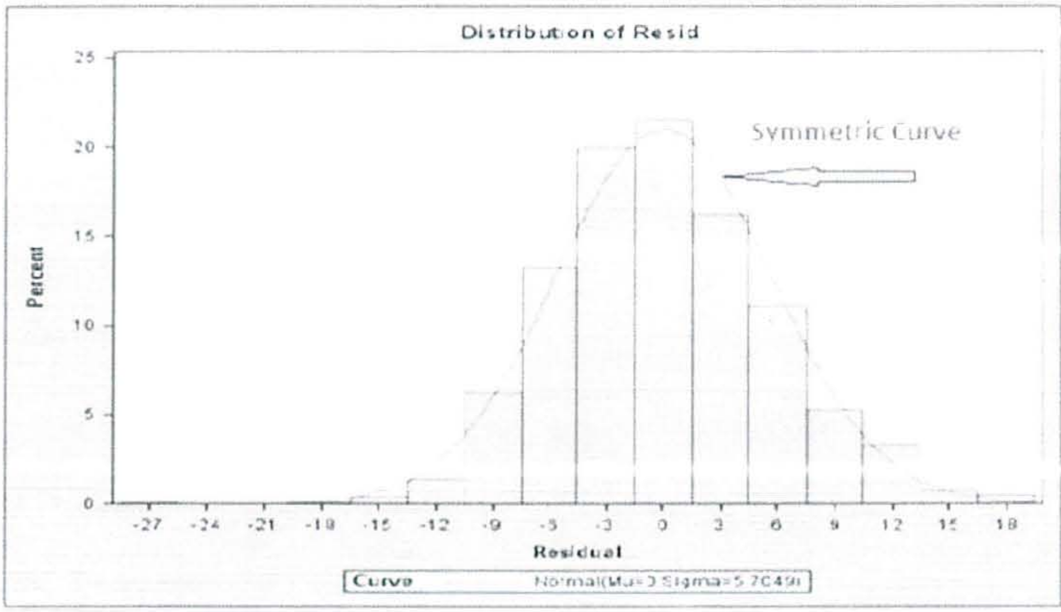

Figure 4: Histogram of mixed effects for residuals.

The histogram of residual sketched in Figure 4 shows that considerably all observations are at the center and the normal curve is symmetrical about the mean of the residuals. It conforms the validity of normality assumptions (mean of residual is zero and the distribution of the data is symmetric around the mean). Finally, we have to consider the scatter plot of standardized residual versus predicted values to detect whether the model is debit or not.

Figure 5 shows that a scatter plots of standardized residual versus the predicted values that includes random effects, $y - X\hat{\beta} - Z\hat{U}$ versus $X\hat{\beta} + Z\hat{U}$, of the response variable (yield of barley). It shows that almost all observations are within the indicated horizontal band (plus or minus two) and the observations are randomly distributed around a horizontal line passing through zero. From this, we can see that constant variance across predicted values and the normality assumptions are fulfilled.

The scatter plot of standardized residual versus predicted values are also used to identify potential outlying and influential observations. Therefore, there is no apparent outlier and influential observations so that it makes us confined with the original data for further analysis and interpretation.

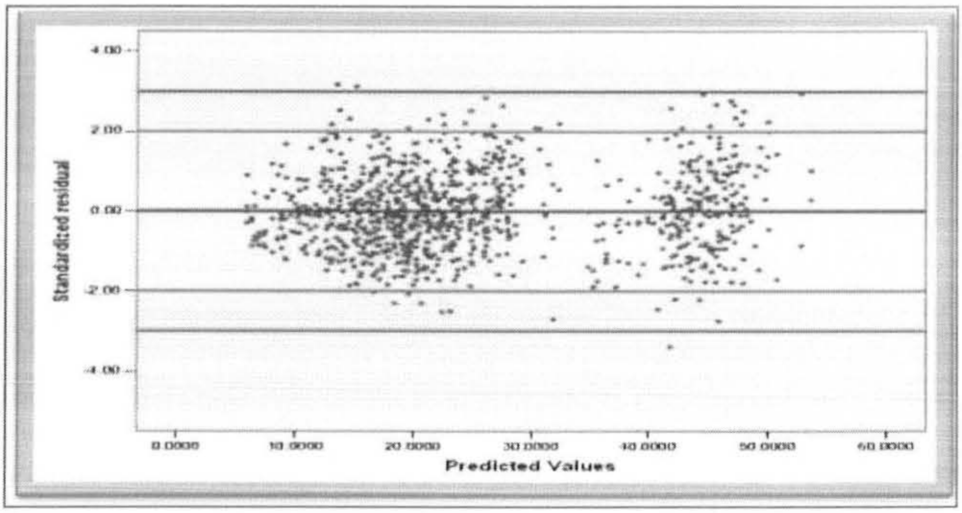

Figure 5: Scatter plot of standardized residuals versus predicted values for mixed effects.

From the three figures (Fig. 3, 4 and 5), Figure 3 shows that the error term and the response variable (yield of barley) have a linear relationship, Figure 4 indicates that the mean of the error term is zero, and finally Figure 5 shows that the variance of the error term is constant. The assumptions of normality fulfilled without transformation of the response variable and uphold the analysis and inference with original data.

All of the above conclusions are checked by PROC UNIVARIATE procedures using SAS for standardized residual of Goodness-of-Fit Tests for Normal Distribution (normality tests) as the p-value (>0.150) of Kolmogorov-Smirnov is greater than level of significant, 0.05 (Appendix C of Summary 2a). The Kolmogorov-Smirnov statistic tests hypothesis that the data are normally distributed and a low significance value (generally less than 0.05) indicates that the distribution of the data differs significantly from a normal distribution (SPSS results coach, Version 13.0). In addition to this the skeweness of the residual (0.09005799- Appendix C of Summary 2a) is between plus and minus one. A measure of skewness is greater or less than one or minus one indicates a distribution that differs significantly from a normal, symmetric distribution (SPSS results coach, Version 13.0).

From the results of model checking, normality assumption is satisfied and the original data is appropriate and can be used in statistical analysis and inference without transformation of the response variable.

To see whether natural logarithm transformation is appropriate for the response variable, the following three figures are plotted after natural logarithm transformation of the response variable.

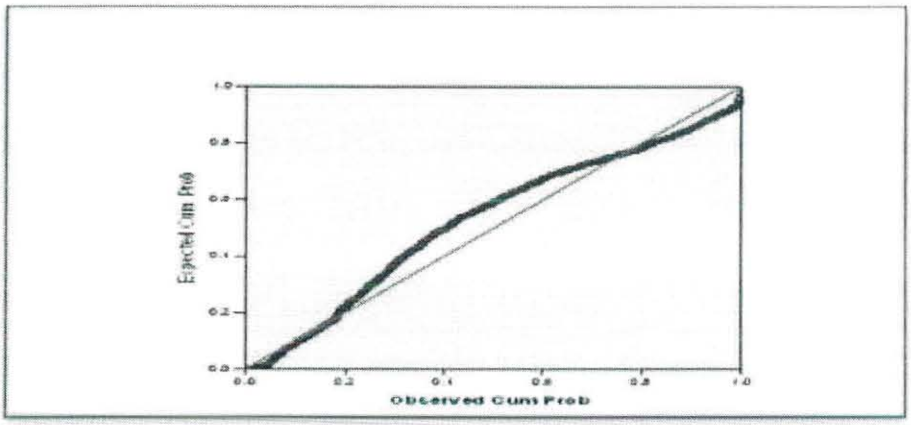

Figure 6: The normal Probability plot for mixed effects of residuals after natural logarithm transformation.

The normal probability plot of Figure 6 does not show a scatter around the straight line and shows that the hypothetical normality assumption fails. It also shows certain patterns which indicate the relationship of the error term with response variable are not linear.

In addition to PP-plot, the histogram of residual (Figure 7) shows that the distribution of the residual appears slightly skewed to the left with a few large negative outliers. This may bias the analysis and inferences for the response variable and the normality assumptions are failed.

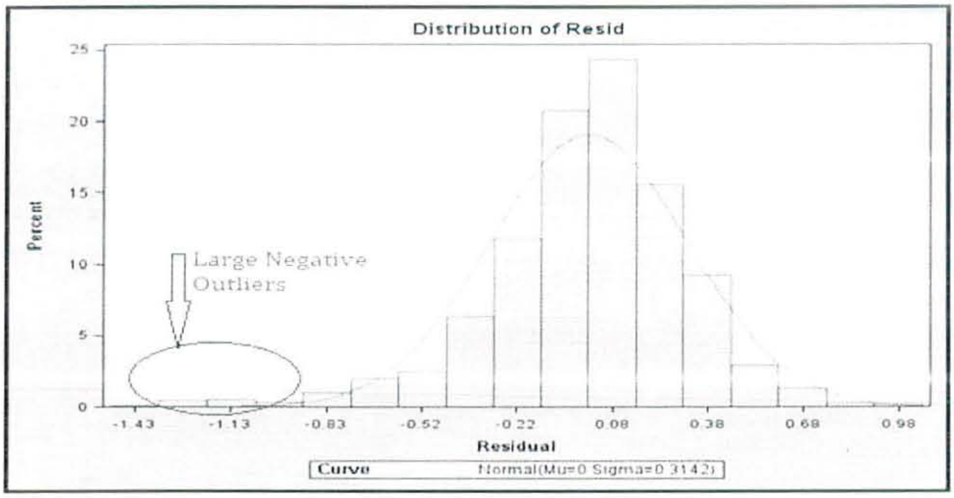

Figure 7: Histogram of residuals for mixed effects after ln transformation

The scatter plot of standardized residual versus predicted values of Figure 8 shows that some observations at the left side fall below minus two. And hence they are not contained in the suggested horizontal bands. In addition to this, the plots are very scattered at the left side when we compare them to those in the right. It has an indication that the variance of error term is not constant. Other transformations like the square root and inverse also show the same characteristics as natural logarithm transformation. The fact that response variable after transformations does not fulfill the normality assumption. Thus, it is better to use original data for further analysis and inference of variance component.

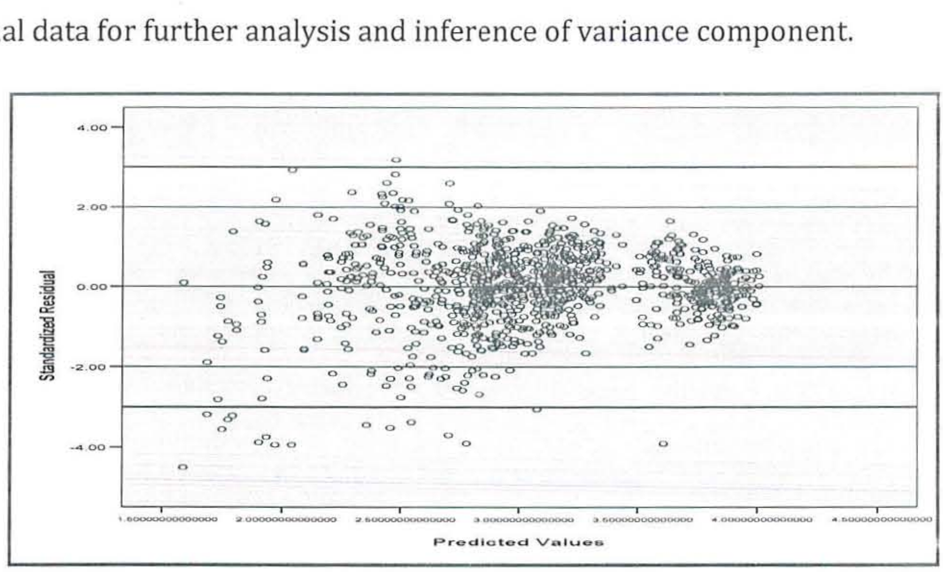

Figure 8: Scatter plot of standard residual versus predicted values for mixed effects after natural logarithm transformation.

CHAPTER FIVE: RESULTS AND DISCUSSION

5.1. Introduction

This chapter provides results of covariance estimation; fixed effect estimation and random effects prediction discussion based on a Henderson method III, ML and REML approach to linear mixed models. This approach simplifies and unifies many common statistical analyses including those involving repeated measures, random effects, and random coefficients. In this thesis, the data analysis is concerned with mixed model effects. The basic assumption in this data analysis is that the response variable is linearly related to unobserved multivariate normal random variables.

5.2. Exploratory Data Analysis

A first step in any statistical data analysis is an exploratory data analysis. In order to get insight about the variables within the data set various exploration techniques were applied. An important goal when exploring the data is choosing an appropriate model (Verbeke and Molenberghs, 2000).

From Appendix C of Summary 3, the overall yield of data ranges from 1.20 to 75.25 quintals per hectare with mean of 25.27 quintals. The mean yield for environment ranged from 16.80 (in environment 4 - SHENO with 150kg of fertilizer) to 44.214 (in environment 2-BEKOJI with 150kg of fertilizer) q/h. A line with the highest yield is EH1700/F7.B1.63.70 (variety 58), which scored the overall maximum yield of 70.25 q/h while line 16, BYDV48/97, scored a minimum yield of 1.2 q/h. Similar to these the overall minimum and overall maximum yield occurs in environment 3 (at SHENO with 100kg of fertilizer) and in environment 1 (at BEKOJI with 100 kg of fertilizer), respectively. These descriptive statistics give insight into the forgoing estimation and inference like the variation of response variable, covariance components of G and R, and predicted means for fixed and random effects.

5.3. Estimation of Variance Components Using Data from Simple Lattice Designs

As stated in chapter three, model fitting exercise consists of three parts; estimating variance parameters, fixed effects, and random effects. A key assumption in the foregoing analysis for the model that **U** and $\boldsymbol{\varepsilon}$ are normally distributed with zero mean vector and covariance matrix of **G** and **R** given in Section 3.3. The estimates of **G** and **R** are calculated using SAS procedure of Henderson method III, ML and REML methods as explained in the review chapter. The fitted model was obtained from model selection procedure as specified in chapter four, i. e., yield as the dependent variable, environment, rep(env) and block(rep) as fixed effects, and variety and the interaction of variety by environment as random effects.

5.3.1. Henderson's Method III

Searle et al (2006, p. 312) summarize different computer packages with their respective procedures to estimate variance parameters and in particular states that the SAS procedure VARCOMP calculates ANOVA estimators based on Henderson's method III from SAS Type I sums of squares. The results are summarized in Table 5.3.1.

The variance component estimated for the random parts of the model, i.e., for variety and variety*environment are 12.08949 and 10.83595, respectively, while the variance component of the error term is 33.51709; it represents the error that remains after the fixed effects and random effects are accounted for. The results of the analysis show that the variance attributable to variety (12.08949) is considerably larger than the variance attributable to variety*environment (10.83595) in magnitude.

Table 5.3.1: Estimation of Variance Components as **β** and **u** are fixed for simple lattice designs.

Type 1 Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Varcomp Estimates
env	7	133219	19031	Var(Error) + 2 Var(Variety*env) + 64 Var(rep(env)) + 128 Var(env)	145.37577
rep(env)	8	3150.719847	393.839981	Var(Error) + 8 Var(Block(rep)) + 64 Var(rep(env))	5.75008
Block(rep)	14	1114.581297	79.612950	Var(Error) + Var(Variety*env) + 8 Var(Variety) + 64 Var(Block(rep))	0.000
Variety	63	14157	224.720382	Var(Error) + 1.7778 Var(Variety*env) + 14.222 Var(Variety)	12.08949
Variety*env	441	24338	55.188992	Var(Error) + 2 Var(Variety*env)	10.83595
Error	490	16423	33.517091	Var(Error)	33.51709
Corrected Total	1023	192403			

This model is fitted considering that there is no difference between **β** and **u** and apply the VARCOMP procedure in SAS. Henderson's method III in SAS does not give any fixed effects estimates or means, it is simply for estimating variance components.

5.3.2. Maximum Likelihood Estimation

The maximum likelihood (ML) technique is used in SAS for estimating the variance components. As discussed in chapter three, estimates of **G** form a diagonal matrix of 10.9331 for the first 64 elements (variety) and a diagonal matrix of 11.0864 for the last 512 elements (variety*environment). An estimate of variance for residuals is 32.1544**I**. Hence, estimates of variance-covariance for the random effects including the error term are given as:

$$\text{var}\begin{bmatrix} u_1 \\ u_2 \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 10.9331\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 11.0864\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 32.1544\mathbf{I} \end{bmatrix}$$

where the **I**'s on the main diagonal along with numbers are the identity matrix of 64x64, 512x512 and 1024x1024 dimensions for the first, second and the third column and u_1 and u_2 represents for variety and variety*environment, respectively.

5.3.3. Residual Maximum Likelihood Estimation

The REML algorithm (using Newton's-Raphson) is used to estimate treatment effects and variance components in a linear mixed model with one fixed effect and two random effects (variety and variety*environment). This algorithm allows an output of the model, variance components, variance-covariance matrix of the estimated components, deviance of the fitted model and Wald tests for all fixed effect terms. The estimated variance components of the random effects (variety and variety*environment) and an error term are shown in the following matrix.

$$\text{var}\begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} 12.0895\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 10.8360\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 33.5171\mathbf{I} \end{bmatrix}$$

Searle et al (2006) pointed out that one of the attractive features of REML, for balanced data, is that it produces exactly the same estimates as VARCOMP (via ANOVA) and this is true in my case as we have seen from Table 5.3.3. This implies Optimal Minimum Variance properties, and it shows that REML estimates in that context do not rely on any normality assumption since only moment assumptions are involved. Table 5.3.3 shows the overall summary estimate for variance components and differences of each other.

Table 5.3.3: Estimation of Variance Component with Different Techniques.

Covariance components	Techniques					
	VARCOMP	ML	REML	REML-VARCOMP	REML-ML	ML-VARCOMP
Variety	12.08949	10.9331	12.0895	1E-05	1.1564	-1.15639
Var*env	10.83595	11.0864	10.8360	5E-05	-0.2504	0.25045
Residual	33.51709	32.1544	33.5171	1E-05	1.3627	-1.36269

The difference between estimates of REML and ML for variety, variety*environment and the error term are 1.1564, -0.2504 and 1.3627 ($\sigma^2_{u_1}$ REML - $\sigma^2_{u_1}$ ML , $\sigma^2_{u_2}$REML - $\sigma^2_{u_2}$ ML and σ^2_{ε}REML - σ^2_{ε} ML), respectively. The reason that ML produces smaller estimate of error term is due to the fact that the ML procedure does not take into account the number of degree of freedom lost when estimating parameters of the model.

The contribution of each random factor in the model variability is determined by the corresponding value of Inter-Class Correlation Coefficient (ICC) of random variables. The larger the inter-class correlation coefficient values of the random variable the larger the contribution of that random variable for the variation of response variable. The ICC values of variety for VARCOMP, ML and REML are 26.51%, 25.37% and 26.51%, respectively. And the ICC values of variety*environment for VARCOMP, ML and REML are 24.43%, 25.63% and 24.43%, respectively. In practice the best approach to estimate the variance components is to work out in all VARCOMP, ML and REML estimation techniques and compare them (as cited by Shauh, R., 2002). If the difference is not too great, then either method can be chosen. If the difference is too great, then one must possibly look at the standard errors of the variance component estimates. Therefore, a difference of ICC with all techniques is too small and estimation based on REML techniques is best to estimate variance components in mixed model for random factors.

Before we proceed to test covariance estimates, it is better to discuss factors that affect the response variable in terms of ($\widehat{G}$) for ML and REML.

Since linear mixed model equations are extended from normal equations, the solutions to these normal equation expressions assumes that $\widehat{\mathbf{G}}$ is nonsingular. Since $\widehat{\mathbf{G}}$ is nonsingular, the inverse exists and hence can obtain unique eigenvalues of $\widehat{\mathbf{G}}$. This eigenvalues indicate the contribution of each random variable in the model.

In the extreme cases, when the eigenvalues of $\widehat{\mathbf{G}}$ is very large, $\widehat{\mathbf{G}}^{-1}$ contributes very little to the model and $\hat{\mathrm{u}}$ gets close to what fixed effects parameters contained in the model. On the other hand, when the eigenvalues of $\widehat{\mathbf{G}}$ are very small, $\widehat{\mathbf{G}}^{-1}$ dominates the model and $\hat{u}$ gets close to 0. For intermediate cases, $\widehat{\mathbf{G}}^{-1}$ can be viewed shrinking the fixed-effect estimates towards 0 (Robinson 1991). According to Robinson 1991, a random variable whose eigenvalue in the range of 10 to 100 is considered as intermediate.

From the two matrices given above, $\widehat{\mathbf{G}}$ is a diagonal matrix in both ML and REML estimation techniques. Beside this, the eigenvalues of $\widehat{\mathbf{G}}$ are the same as those of the elements of the matrix of $\widehat{\mathbf{G}}$ on the main diagonal since the solution of the characteristics equation given by $|\hat{G} - \lambda I| = 0$ gives the same results as that of $\widehat{\mathbf{G}}$ elements. Hence, the values of the eigenvalues for the first 64 and the last 512 elements in ML and REML are 10.9331 and 11.0864 and 12.0895 and 10.8667, respectively.

The inverse of $\widehat{\mathbf{G}}$ is simply the inverse of those elements on the main diagonal. As a result, the inverse of $\widehat{\mathbf{G}}$ are 0.091465**I** and 0.090201**I** to the first 64 elements (variety) and the last 512 elements (variety*environment) using ML respectively. Similarly the inverse of $\widehat{\mathbf{G}}$ are 0.082716**I** and 0.092288**I** to the first 64 elements (variety) and the last 512 elements (variety*environment) using REML respectively where **I** is an identity matrix with dimension of 64x64 and 512x512 for the first and the last diagonal matrices. Therefore, as the Eigenvalues of $\widehat{\mathbf{G}}$ are intermediate (both variety and variety*environment), $\widehat{\mathbf{G}}^{-1}$ is shown shrinking the fixed-effects estimates towards 0. This implies that the inclusion of random effects play a very significant role in explaining the variation of response variable in both techniques.

5.4. Test Statistics of Variance Estimates and Inference for Simple Lattice Designs

In a linear mixed model, variance component estimation produces point estimates of each parameter. These point estimates are valuable in making inferences. For inferences concerning covariance parameters in linear mixed model, it is possible to use likelihood-based statistics. One common likelihood-based statistic is *Wald statistic*. It is computed as the parameter estimate divided by its *asymptotic standard error*. The asymptotic standard errors are obtained from the inverse of second derivative matrix of the likelihood with respect to each of the covariance parameters. The Wald *statistic* is valid for large samples, but it can be unreliable for small data sets and for parameters such as variance components, which are known to have a skewed or bounded sampling distribution (SAS manual 1999). The observed Fisher information matrix is evaluated at the final iteration covariance parameter estimate. Therefore, PROC MIXED procedure uses these asymptotic variance-covariance estimates by profiling out the residual variance from the likelihood to calculate different statistic, like Wald statistic.

Table 5.4: Covariance Parameter Estimates for SLD together with Wald Statistic.

Covariance Parameter Estimates using ML					Covariance Parameter Estimates using REML				
Cov Parm	Estimate	Standard Error	Z Value	Pr > Z	Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
Variety	10.9331	2.5808	4.24	<.0001	variety	12.0895	2.7395	4.24	<.0001
Var*env	11.0864	2.0755	5.34	<.0001	var*env	10.8360	2.1427	5.07	<.0001
Resid	32.1544	2.0135	15.97	<.0001	Residual	33.5171	2.1333	15.68	<.0001

Table 5.4 displays covariance parameter estimates together with asymptotic standard errors and the Wald statistic of SLD using ML and REML techniques. As we have seen from Table 5.4, the Wald test statistic indicates that both random variables (variety and variety*environment) and residuals are highly significant (p-value <0.0001) for ML and REML methods. As the p-values of the covariance components are small, i. e., <0.05, (Table 5.4) the random variables are useful in the model to minimize error mean squares of

response variable. It is because that the error mean square of response variable is the totality of both experimental and random effects.

In relation to Wald test statistic, a better alternative is the chi-square likelihood ratio test i.e. χ^2. This is commonly used when we are testing that a variance component equals or not to its lower boundary constraint of zero (Self and Liang 1987). It ensures that whether using random effect really increases the variance of response variable or not. If the test statistic shows that the variance is not significantly different from zero, then the inclusion of the random variable is meaningless. But, if the test statistic indicates that the variance is above the lower boundary constraints, then the variance of the response variable also increases and this in turn reduces the overall estimation, confidence interval, prediction interval and test statistic is based on this variance. The square of Wald statistic is approximately chi-square distribution with one degree of freedom for that Wald test statistic is approximately normal. Based on Table 5.4, the chi-square statistic for variety and variety*environment are 17.9776 and 28.5156 for ML techniques and 17.9776 and 25.5625 for REML methods respectively. The 95 and 99 percent of the chi-square distribution with one degree of freedom have critical values of 3.84 and 5.02, respectively. Therefore, the test statistic with 1% indicates that both random variables are statistically significant, meaning that the estimate of the variance of all random effects are above the lower boundary constraint of zero. This implies that each random factor has a contribution for increasing variance of response variable and in turn minimizes error mean squares of response variable.

5.5. Variance Components Estimation for Simple Lattice Design with Missing Plots

Once trials are designed as simple lattice, problems such as water losing or damage of some plots due to rodent or birds may still increase the non-balance between treatments and blocks which exists in such designs. Missing plots due to those problems affects performance of some specific treatment which becomes victim of the scenario. This is a very common occurrence in agricultural research field trials and requires investigation in order to provide advice in this regard. We therefore purposely omitted the yield from twenty nine plots randomly from both replications. The result shows variance component for simple lattice design with missing plots using Henderson's method III, ML and REML techniques.

Table 5.5 shows that the results of variance component estimates of simple lattice design with missing plots using VARCOMP (ANOVA) techniques, i.e., for variety and variety*environment are 12.39311 and 10.74544 respectively; and the variance component of the error term is 33.42304 which represents the error that remains after the fixed effects and random effects are accounted for.

Table 5.5: Estimation of Variance Components with **β** and **u** has no difference for SLD with missing plots.

Type 1 Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Varcomp Estimates
env	7	127761	18252	Var(Error) + 1.996 Var(Variety*env) + 0.0468 Var(Variety) + 0.0194 Var(Block(rep)) + 62.251 Var(rep(env)) + 124.5 Var(env)	143.29931
rep(env)	8	3048.621346	381.077668	Var(Error) + 0.004 Var(Variety*env) + 0.004 Var(Variety) + 7.8022 Var(Block(rep)) + 62.248 Var(rep(env))	5.71236
Block(rep)	14	1078.790194	77.056442	Var(Error) + Var(Variety*env) + 7.8156 Var(Variety) + 62.226 Var(Block(rep))	0.000
Variety	63	14096	223.744052	Var(Error) + 1.7741 Var(Variety*env) + 13.819 Var(Variety)	12.39311
Variety*env	429	23518	54.819503	Var(Error) + 1.9912 Var(Variety*env)	10.74544
Error	474	15843	33.423044	Var(Error)	33.42304
Corrected Total	995	185345			

The maximum likelihood estimation of variance component for SLD with missing plots is like that of SLD. The main diagonal matrix of **G** for the first 64 elements (variety) and the last 499 elements (variety*environment) are 11.2295 and 11.0243, respectively. The estimate of variance for residuals is 32.0152**I**. Hence, estimates of variance-covariance for the random effects including the error term for SLD with missing plots is given as:

$$\mathrm{var}\begin{bmatrix} u_1 \\ u_2 \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 11.2295\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 11.0243\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 32.0152\mathbf{I} \end{bmatrix}$$

where the **I**'s on the main diagonal along with numbers are the identity matrix of 64x64, 499x499 and 995x995 dimensions for the first, second and third column and $\mathbf{u}_1$ and $\mathbf{u}_2$ represents for variety and variety*environment, respectively.

The REML algorithm (Newton-Raphson) techniques allow us estimation of variance components for SLD missing plots. The estimated variance components of the random effects (variety and variety*environment) and an error term for SLD missing plots based on REML techniques are shown in the following matrix.

$$\mathrm{var}\begin{bmatrix} u_1 \\ u_2 \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 11.9216\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 10.7974\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 33.3510\mathbf{I} \end{bmatrix}$$

The difference between estimates of REML and ML for variety and variety*environment are 0.6921 and -0.2296 ($\sigma^2_{u_1}$ REML - $\sigma^2_{u_1}$ ML and $\sigma^2_{u_2}$REML - $\sigma^2_{u_2}$ ML), respectively. There is a large difference of error term estimates between REML and ML techniques (σ^2_{ε}REML - σ^2_{ε}ML =1.3358). The difference between estimates of ML and VARCOMP methods for variety, variety*environment and error term are -1.16361, 0.27886 and -1.40784 ($\sigma^2_{u_1}$ ML - $\sigma^2_{u_1}$ VARCOMP $\sigma^2_{u_2}$ML - $\sigma^2_{u_2}$ VARCOMP and σ^2_{ε}ML - σ^2_{ε}VARCOMP), respectively. The difference between REML and VARCOMP estimates of variance components for SLD with missing plots are not zero. The difference estimates of variance component between REML and VARCOMP techniques for random factors and error term are -0.47151, 0.05196 and -0.07204, respectively. From the differences, the one being positive (for $\sigma^2_{u_2}$) and the two being negative (for $\sigma^2_{u_1}$ *and* σ^2_{ε}) we can conclude that σ^2_{ε} benefit from REML and this shows that the REML technique minimizes the error mean squares and increases the variation of the response variable.

5.6. Test Statistics of Variance Estimates and Inference to Simple Lattice Design with Missing Plots

Table 5.6 shows covariance parameter estimates together with asymptotic standard errors and the Wald statistic of SLD with missing plots using ML and REML techniques. As we have seen from table 5.6, the Wald test statistic indicates that both random variables (variety and variety*environment) and residuals are highly significant (p-value <0.0001) for ML and REML techniques.

Based on Table 5.6, the chi-square statistic of variety and variety*environment for ML and REML techniques are 17.8929 and 24.7009 and 17.8929 and 27.5625, respectively. The 95 and 99 percent of the chi-square distribution with one degree of freedom have critical values of 3.84 and 5.02, respectively. Therefore, the test statistic with 1% indicates that both random variables are statistically significant, meaning that estimates of the variance of all random effects are above the lower boundary constraint of zero like that of SLD.

Table 5.6: Covariance parameter estimates for SLD with missing plots together with Wald statistic.

Covariance Parameter Estimates using ML				
Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
Variety	11.2295	2.6563	4.23	<.0001
Var*env	11.0243	2.1013	5.25	<.0001
Resid	32.0152	2.0356	15.73	<.0001

Covariance Parameter Estimates using REML				
Cov Parm	Estimate	Standard Error	Z Value	Pr > Z
variety	11.9216	2.8210	4.23	<.0001
Var*env	10.7974	2.1714	4.97	<.0001
Residual	33.3510	2.1604	15.44	<.0001

Note that the test statistic of variance component estimates and its inference for SLD and SLD with missing plots are almost the same in this study. This is because that it uses the asymptotic standard errors and that the numbers of observations in both SLD and SLD with missing plots are asymptotically large.

5.7. Comparative Advantage of Mixed Effects Model over Fixed Effects Model.

As it is generally accepted, General Linear Model (GLM) is a regression model and does not include the random effect except the random error terms of the model. Hence, there is no variance due to the random effect other than the variance of random error terms when calculating the variances (standard error) of the response variable, estimates of the parameters, testing and constructing confidence interval. As a result, all variances expected from the model, if any, are included in the variance of the disturbance term inappropriately since all factors in GLM are fixed effects which do not endure the inclusion random effects. In such kinds of models the variance of the error terms boost very much. This illustrates the danger of misusing the statistical models and the rush to irrelevant conclusion.

However, there are random factors included in the mixed model that contributed their lion share in estimating the variance of the response variable. The inclusion of the random effects in the model removes the downward bias of variance of the response variable and the boosted variance of the error terms of GLM.

To see the advantage of mixed model over GLM, we need to observe the standard error of mixed model and GLM. The standard errors of residuals from mixed model for SLD are 2.0135 and 2.133075 in ML and REML techniques (Table 5.4). Similarly, the standard errors of residuals from mixed model considering SLD with missing plots for ML and REML methods are 2.0356 and 2.1604, respectively, while the standard errors of residuals from GLM for SLD is 5.789395 and the standard errors of residual from GLM for SLD with missing plots is 5.781267 (Summary 4). The ratios of standard errors of GLM to that of mixed model using SLD is about 2.714086 for both ML and REML; whereas, the ratios of SLD with missing plots for ML and REML techniques is about 2.676017. From this, we conclude that the standard error of GLM is approximately 2.71 times that of mixed model in SLD and 2.67 in SLD with missing plots. These results tell us that a mixed model is more valuable than a GLM to remove the downward bias of variance of the response variable and the boosted variance of the error terms of GLM.

5.8. Statistical Properties of Fixed & Random Effects in Mixed Model and Estimation for Simple Lattice Design

As described in chapter three, the standard method is to solve the mixed model normal equations to obtain estimates of $\boldsymbol{\beta}$ and $\mathbf{u}$. When the inverse of $(\mathbf{X'X})$ does not exist, a generalized inverse can be used in its place. But, in our case $\mathbf{X}$ has no full column rank and hence used a generalized inverse of $\mathbf{X}$. Because, the sum of the last eight columns of the design matrix for fixed effect gives the first column which represents the intercept. As a result, the true inverse of $X' \widehat{V}^{-1}X$ does not exist. Hence, generalized inverse is used instead. Therefore, to obtain the fixed effect estimates in both method of estimation, we have to use the generalized inverse of $X' \widehat{V}^{-1}X$ to overcome the nonexistent of inverse.

Inferences about fixed effects have come to be called *estimates*, whereas those that concern random effects are known as *predictions*. If $\mathbf{G}$ and $\mathbf{R}$ are known, $\widehat{\boldsymbol{\beta}}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and $\widehat{\mathbf{u}}$ is the best linear unbiased predictor (BLUP) of $\mathbf{u}$. They are *best* in the sense that they minimize the sampling variance, *linear* in the sense that they are linear functions of the response variable and *unbiased* in the sense that $\mathbf{E[BLUE(\boldsymbol{\beta})] = \boldsymbol{\beta}}$ and $\mathbf{E[BLUP(U)] = E(U)}$ i.e., $\mathbf{E(\widehat{U}) = E(U)}$. However, $\mathbf{G}$ and $\mathbf{R}$ are usually unknown and are estimated using one of the aforementioned (VARCOMP, ML and REML) methods. These estimates, $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{R}}$, are therefore simply substituted into the preceding expression to obtain the approximate variance-covariance matrix of $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \widehat{\mathbf{u}} - \mathbf{u})$. In this case, the BLUE and BLUP acronyms are no longer applied, but the word empirical is often added to indicate such an approximation. The appropriate acronyms thus become EBLUE and EBLUP (McLean and Sanders 1988). As a result, the analysis and inferences using ML and REML estimation for EBLUE and EBLUP are the same.

As we have seen from Table 5.8A, the PROC MIXED procedure in SAS sets zero the estimates of environment eight. However, these are not the actual estimates of those factors. In mixed model, if the model under consideration is a no-intercept model and the variable under consideration has different levels, PROC MIXED in SAS sets the estimates of each level of the variable by subtracting the estimates of the last level from the estimates of the remaining levels of the variable. If the model is intercept model, the estimates of each variable is the sum of intercept and the estimate set for specific level of the variable under consideration (SAS 1999). Having this in mind, Table 5.8A shows the difference mean effect yield of barley between environment 8 (North Gondar with 150 kg of fertilizer) and specific environment level. Therefore, the difference mean effect yield of barley between environment 8 (North Gondar with 150 kg of fertilizer) and level one (BEKOJI with 100 kg of fertilizer) is 20.8601 and so on. Table 5.8A also shows that the difference estimated mean effects between North Gondar with 150 kg of fertilizer and all other levels are significantly different from zero at alpha value greater than 0.0001 (as p-value is <0.0001). But, the difference predicted mean effects between North Gondar with 150 kg of fertilizer and HOLETTA with 150 kg of fertilizer and SHENO with 100 kg of fertilizer are not statistically different from zero at alpha value of 0.05.

Table 5.8A: Empirical Best Linear Unbiased Estimates of Fixed Effects (Environment).

Solution for Fixed Effects (environment)						
Effect	env	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		22.51	1.3077	63	17.21	<.0001
env	1	21.188	1.1624	441	18.23	<.0001
env	2	21.743	1.1624	441	18.7	<.0001
env	3	-13.813	1.1624	441	-11.88	<.0001
env	4	-5.3248	1.1624	441	-4.58	<.0001
env	5	1.6869	1.1624	441	1.45	0.1475
env	6	3.8564	1.1624	441	3.32	0.001
env	7	-7.2811	1.1624	441	-6.26	<.0001
env	8	0	.	.	.	.

Table 5.8B: Least Squares Means Estimates for Fixed Effects.

Least Squares Means						
Effect	env	Estimate	Standard Error	DF	t Value	Pr > |t|
env	1	43.4088	0.7774	764	55.84	<.0001
env	2	44.2138	0.7774	764	56.87	<.0001
env	3	10.7645	0.7774	764	13.85	<.0001
env	4	16.8012	0.7774	764	21.61	<.0001
env	5	22.7047	0.7774	764	29.20	<.0001
env	6	23.8516	0.7774	764	30.68	<.0001
env	7	17.8699	0.7774	764	22.99	<.0001
env	8	22.5488	0.7774	764	29.00	<.0001

Table 5.8B shows actual estimated means for fixed effects. It is obtained by summing estimates of the intercept (=environment 8) and estimates of respective fixed effect levels. For instance, the actual estimated mean effect for North Gondar with 150 kg of fertilizer is 22.5488+0 (=22.5488) and the actual estimated mean effect for BEKOJI with 100 kg of fertilizer is 22.5488+20.8601 (=43.4088).

The predicted mean for environment 3 is 10.7645q/h which is the smallest predicted mean among all environments. The predicted mean for the second environment level is 44.2138 quintals per hectare that shows the highest predicted mean. The second highest predicted mean has been seen at environment 1 (=43.4088 q/h). This tells us BEKOJI has the highest predicted mean with both fertilizer doses. This result logically agrees with the idea of different ecologists pointing that climate of BEKOJI is suitable for barley production. Table 5.8B also shows that all predicted means for fixed effects have p-values less than 0.0001 and were statistically different from zero. This indicates that all fixed effect levels have different mean effects to barley production.

The empirical estimates of random effects with approximate standard errors, the t-test statistic, and p-value are given in Appendix C of Table C3 and C4. Empirically best linear unbiased predictor (EBLUP) of yield in Table C3 shows that the overall mean plus average effect of each variety to the mean yield of barley. These is true for variety*environment as shows in Table C4.

In linear mixed models, prediction means of random effects should be used as a selection index based on the degree of estimated predictors to the response variable without bothering about its significance (Temesgen, 2009/10). This is because the standard errors are not obtained from the covariance matrix of empirical estimates ($\hat{u}$). The standard errors are rather obtained from the covariance matrix of the difference between empirical estimates and the unknown random parameters ($\hat{u}$ - u) (Verbeke and Molenberghs, 2000). So, the higher the empirical estimated predictor of random effects the larger contribution to mean yield. Therefore, Table 5.8C shows that the top two varieties have best mean effects to barley production when we compare all 64 varieties each other and each environment has two top varieties having best mean effects. The overall empirical estimated predictors of random effects are illustrated in Appendix C of Table C3 and C4.

As we see in Table 5.8C, varieties EH1700/F7.B1.63.70, EH1643/F7.B1.16.27.25 and EH1748/F7.B1.111.85 have the highest mean effects on the mean yield of barley in decreasing order. But, when we compare each variety with a specific environment, line IWFBSO 76/97 and line EH1671/F7.B2.34.41 have best mean effects in BEKOJI with 100 and 150 kg fertilizer doses, respectively. Variety EH1607/F6.3H.5.8.3 using 100 kg of fertilizer and variety BYDV42/97 with 150 kg of fertilizer have produced the highest mean yield effects in SHENO. In HOLETTA, variety EH1642/F7.B1.15.25.23 produces the highest mean effects to yield using 100 kg of fertilizer and variety EH1668/F7.B1.31.36 produces the highest mean yield effects with 150 kg of fertilizer. The unique result in this study is that the local check with fertilizer doses of 150 kg produces the highest mean effects to yield in North Gondar.

Table 5.8C: Selected Empirical Best Unbiased Linear Predictions for random effects.

Effect	var	env	Estimate	Std Err Pred	DF	t Value	Pr > |t|
variety	58		6.0353	1.6464	512	3.67	0.0003
variety	50		5.0713	1.6464	512	3.08	0.0022
var*env	8	1	4.2291	2.3581	512	1.79	0.0735
var*env	42	1	3.864	2.3581	512	1.64	0.1019
var*env	55	2	3.4031	2.3581	512	1.44	0.1496
var*env	11	2	2.8346	2.3581	512	1.2	0.2299
var*env	20	3	2.8218	2.3581	512	1.2	0.232
var*env	51	3	2.4474	2.3581	512	1.04	0.2998
var*env	14	4	3.5503	2.3581	512	1.51	0.1328
var*env	46	4	3.1019	2.3581	512	1.32	0.189
var*env	49	5	4.0785	2.3581	512	1.73	0.0843
var*env	52	5	3.3532	2.3581	512	1.42	0.1556
var*env	54	6	3.0048	2.3581	512	1.27	0.2032
var*env	27	6	2.9776	2.3581	512	1.26	0.2073
var*env	48	7	1.9727	2.3581	512	0.84	0.4032
var*env	17	7	1.8693	2.3581	512	0.79	0.4283
var*env	64	8	2.539	2.3581	512	1.08	0.2821
var*env	8	8	1.9171	2.3581	512	0.81	0.4166

CHAPTER SIX: CONCLUSION AND RECOMMENDATIONS

6.1. Conclusion

The estimation of parameters in mixed linear model requires the specification of **X**, **Z**, **G**, and **R** even though **X** and **Z** have their own specific form and known elements. This is because that knowledge of the structure of G and R helps to estimate the variance of the random factors. There are many different covariance structures for **G** and **R** from which one or many others might be reasonable. But, a few brief comments and references are in order. First, subject matter considerations and the objectives stated to undertake analysis; second, when the data themselves are looked to for guidance (an indication), and finally, a likelihood-based approach to the mixed model provides several statistical measures for model adequacy as well. The most commonly used and forwarded structure of **G** and **R** is the one whose matrix is diagonal.

Since linear mixed model equations are extended normal equations, the solutions to these normal equation expressions are assumed to be nonsingular. As a result, it is more useful to use the Newton-Raphson Algorithm which is the second derivative matrix of the objective function evaluated at the optimal point up on attainment. Because, it simplifies the test and confidence interval calculations that are based on asymptotic normality for normally distributed large sample sizes.

Moreover, the convergence criteria of the model are the primary point that were raised in linear mixed model as parameter estimation and the test statistic of the model is based on these convergence criteria. If convergence is met, then the fixed effect estimates are unbiased and the estimates of the variance parameters also tend to be asymptotically efficient. The linear mixed model equation is solved analytically for the optimal value of residual variance using likelihood. This is because it can improve convergence properties and also used to obtain an asymptotic estimator for variance components that helps one in estimating and testing the unknown coefficients of the model.

Mixed model is more valuable than GLM to remove the downward bias of variance of the response variable and the boosted variance of the error terms of GLM. This is because the standard error of GLM is about 2.7 times that of mixed model for both SLD and SLD with missing plots. As we consider mixed model, the contribution of each random factor in the model variability is determined by the corresponding Inter-Class Correlation Coefficient (ICC) of random variables. This means that the larger ICC of random factors the higher the contributions to the variations of the response variable. These factors are imperative in explaining the response variable when used as random factors than fixed effect factors. ICC showed that the contribution of both variety and variety*environment maximizes the variations of the response variable and minimize the error mean squares.

For SLD mixed models, the REML estimation for variance components are indistinguishable from classical estimates obtained from solving the normal equations which set mean squares equal to their expectations. This implies optimal minimum variance properties and REML estimates in that context do not rely on any normality assumption since only moment assumptions are involved. But, for SLD with missing plots of mixed models, the REML estimation for variance components is different from classical estimates obtained from solving the normal equations which set mean squares equal to their expectations. This difference shows that estimation of variance components benefits from REML but VARCOMP does not.

Every method has its own advantages and disadvantages. VARCOMP estimation does not consider the factors as random effect. ML estimation includes factors as the random effect but does not take into account the number of degrees of freedom with the fixed effects of the model. REML does take into account the number of degrees of freedom for the fixed effects of the model. In practice, the best approach to estimate the variance components is to work out in all VARCOMP, ML and REML estimation techniques and compare them. If the difference is not too large then either method can be chosen. If the difference is too large, then one must possibly look at the standard errors of the variance component estimates. Therefore, the differences of variance component estimates between all estimation techniques are too small and REML is the best among all techniques. This implies that REML is the best method of estimation for variance components for SLD and/ or SLD with missing plots.

The predicted mean of each environment (combinations of location by fertilizer) were statistically different and hence all had been contributing different mean effects on yield. With both fertilizer combinations, the highest mean grain yield was observed at BEKOJI, while the smallest mean grain yield was registered at SHENO. From the results, we can conclude that BEKOJI is suitable for barley production with both fertilizer doses (100kg and 150 kg) unlike that of SHENO.

When we consider the random effects of linear mixed models, we should bother about the degree of predicted means of each random factor to the response variable rather than their significance. Therefore, varieties EH1700/F7.B1.63.70, EH1643/F7.B1.16.27.25, EH1748/F7.B1.111.85, and EH1624/F7.1H.2.99 have the highest mean effects on yield of barley, while the local check has the 9th best mean effects. But, when we compare a variety with specific a environment, IWFBSO 76/97 and EH1671/F7.B2.34.4 using 100 kg and 150 kg fertilizer doses respectively, have produced best mean effects at BEKOJI and throughout the country that have similar agro-ecological zones. Variety EH1607/F6.3H.5.8.3 with 100 kg of fertilizer and variety BYDV42/97 using 150 kg fertilizer doses have brought highest mean yield effects in SHENO and other areas of Ethiopia having the same climate. Variety EH1642/F7.B1.15.25.23 and variety EH1668/F7.B1.31.36 have shown the highest mean effects to yield in HOLETTA and other similar areas of the country when 100 and 150 kg of fertilizer is employed, respectively. Variety EH1628/F7.B1.6.22.19 with fertilizer does of 100 kg has the highest mean effect to production of barley in North Gondar and other places of Ethiopia having the same agro-ecological zones. The same is true for local check when we use 150 kg of fertilizer in North Gondar. This indicates that local check should be replaced by other lines which have best mean effects except in North Gondar.

6.2. Recommendations

Based upon the major findings of this paper, the author would like to recommend the following major points to the problems addressed by the study.

- ✓ Barley researchers should use mixed models rather than fixed effect models to minimize the severely inflated experimental error mean square and to attain an acceptable level of experimental precision as the standard error of GLM is higher than that of mixed model.
- ✓ It is recommended that the approach to estimate the variance components in linear mixed model for SLD and SLD with missing plots to be REML because it does take into account the number of degrees of freedom for the fixed effects of the model and does not rely on the assumptions of normality for balanced data.
- ✓ The National Barley Research Program (NBRP) should make variety 8 and 42 available to farmers residing in similar agro-ecological zones to BEKOJI and variety 20 and 14 to locations similar to SHENO. In addition to this, variety 49 and 54 are recommended to similar ecologic zones to HOLETTA. Finally variety 48 and Local check can be adapted in areas having similar agro-ecological zones to North Gondar.
- ✓ NBRP should teach farmers about the technical and field application of newly released barley varieties to increase their barley production.

REFERENCES

1. Akaike, H. (1974). A new look at the statistical model of identification. *IEEE Transaction on Automatic Control, 19,* 716–723.
2. Beckman, P.J., Nachtsheim, C.J., and Cook, R. D. (1987). Diagnostics for mixed-model analysis of variance. *Technometrics,* 29, 413-426.
3. Belsley, D.A., Kuh, E., and Welsch, R.E. (1980). Regression Diagnostics; Identifying Influential Data and Sources of Collinearity, New York: John Wiley & Sons.
4. Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC), the general theory and its analytical extensions. *Psychometrika* 52:345–370.
5. Cavatassi, K., Richards, M.C., and Heppel, V. (2006). Cereal varieties for the organic and low input grower. p. 147–155.
6. Capper, P.M., Entz, M.H., Guilford, R., and Gulden, R. (1988). Crop yield and soil nutrient status on 14 organic farms in eastern portions of the northern Great Plains. Can. J. Plant Sci. 81:351–354.
7. Central Statistical Agency, 2008/09 Agricultural Sample Surveys: Report on Area and Production for major crops statistical bulletin. Addis Ababa Ethiopia.
8. Chow, S.C. and Shao, J. (1988). A New procedure for the estimation of variance components. *Statist. Prof. Letter 6(5)*: 349-355.
9. Christensen, R., Pearson, L. M., Johnson, W. (1992). Case deletion diagnostics for mixed models. *Technometrics* 34:38–45.
10. Crump, S.L. (1951). The estimation of variance components in analysis of variances. *Biometrics Bull.,* 2, 7-11.
11. Cochran, W.F. (1939). The use of analysis of variance in enumeration by sampling. J. *Amer. Stat. Assoc.,* 34, 492-510.
12. Cook, R.D. (1977). Detection of influential observations in linear regression. *Technometrics,* 19, 15-19.
13. Cook, R.D. and Weisberg, S. (1982), Residuals and Influence in Regression, New York: Chapman and Hall.
14. Daniel H.E. (1939). The estimation of components of variance. *J. R. stat. Soc. Suppl.,* 6, 186-197.

15. Duchateau, L. and Janssen, P. and Rowlands, J.G. (1998). "Linear Mixed Models: an Introduction with applications in Veterinary research", ILRI (international Livestock Research Institute) Nairobi, Kenya.
16. Douglas, L., Bera, A. K., Jarque, C. M., and Lee, L. F. (1991), Testing the Normality Assumption in Limited Dependent Variable Models, *International Economic Review*, 25, 563–578.
17. Ellis, R. P., Slafe, G.A., Molina-Cano, L. J., Savin, R., Araus, J.L., and Romagosa, I. (2002). Wilt barley as a source of genes for crop improvement. *Barley Science: Recent Advances from Moleculer Biology to Agronomy of Yield and Quality* pp. 65-67. Haworth Press, Binghamton.
18. Ministry of Water Resources (2001): Initial National Communication of Ethiopia to the United Nations Framework Convention on Climate Change. *Federal Democratic Republic of Ethiopia.*
19. Federal Democratic Republic of Ethiopia (1997): *Environmental Policy.*
20. Fisher, R.A. (1925). Statistical Methods for Research Workers, 1st ed. Oliver & Boyd, Edinburgh and London.
21. Ganguli, M. (1941). A note on nested sampling. *Sankhya 5*, 449-452.
22. Girma, T. (2005). "Estimation of Optimum Plot Dimension and Replication number for Wheat Experiment in Ethiopia" *African Crop Science Journal Volume 1* (pp 11-24).
23. Gurka, M. J. (2006). Selecting the best linear mixed model under REML. *The American Statistician* 60:19–26.
24. Hailu, G.M. (2002). Bread wheat breeding and genetic research in Ethiopia: A Historical perspective. Hailu Gebre-Mariam, Tanner, D.G. and Mengistu Hulluka (Eds.), pp. 73-93. IAR/CIMMY, Addis Ababa, Ethiopia.
25. Harville, D.A. (1977). Maximum Likelihood approaches to variance component estimation and to related problems. *J. Amer. Stat. Ass.* 72: 320–338.
26. Hartley, H. O. and Rao, C.R. (1967). Maximum Likelihood Estimation for the Mixed Model Analysis of Variance Model. *Biometrika vol 54*. NO1/2, PP 93-108.
27. Henderson, C.R. (1984). Application of Linear Models in Animal Breeding, University of Guelph.

28. Henderson, C.R. (1953). Estimation of variance and covariance components. *Biometrics* 9: 226–252.

29. Henderson, C.R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding* (W.D. Hanson and H.F. Robinson, eds.), 141-163. National Academy of Sciences and National Research Council Publication No. 982, Washington, D.C.

30. Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31. 423-447.

31. Hocking, R. R., Green, J. W., and Bremer, R. H. (1989). Variance component estimation with model based diagnostics. *Technometrics* 31:227–239.

32. Hurvich, C. M. and Tsai, C.-L. (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika,* 76, 297-307.

33. ICPPG/FAO.1997. States of the world's plants genetic resources for food and agriculture, Rome, p.185.

34. Jennrich, R.I. and Schluchter, M.D. (1986). "Unbalanced Repeated-Measures Models with Structured Covariance Matrices," *Biometrics,* 42, 805-820.

35. Kackar, R., and Harville, D. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, *79*, 853–862.

36. Kearsey, M. J. Callow, J.A., Ford-Lloyd, B.V., and Newbury, H.J. (1997). Genetic resources and plant breeding. *Biotechnology and Plant Genetic Resources* p, 175, CAB International, New York.

37. Klaus, H. and Kempthorne, O. (2005). Design and Analysis of Experiments. Volume 2: Advanced Experimental Design. ISBN 0-471-55177-5.

38. Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* 38, 963-974.

39. Lindsey, J.K. (1993), *Models for Repeated Measurements,* Oxford: Clarendon Press.

40. Lesaffre, E., Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics* 54:570–582.

41. Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger,R. D. (1996). SAS System for Mixed Models, Cary, NC: SAS Institute Inc.

42. Longford, N.T. (1993). Random coefficient models, 2nd ed. New York: Oxford University Press.
43. McLean, R.A. and Sanders, W.L. (1988). Approximating Degrees of Freedom for Standard Errors in Mixed Linear Models, Proceedings of the statistical Computing Section, *American statistician*, 45, 54-64.
44. Ministry of Trade and Industry (2008). Export hard currency earnings accomplishment report for fiscal years 1999 and 2000. Addis Ababa, Ethiopia.
45. Montgomery, D. C. (1991). Design and Analysis of Experiments. 3th ed, John Wiley & Sons, Inc.
46. Mora, F. and Arnhold, E. (2006). Application of the Bayesian inference and mixed linear model method to maize breeding. Cien. Inv. Agr. (In English) 33(3):185-190.
47. Neyman, J., Iwaszkiewicz, K., and Kolodziejcyk, S.T. (1935). Statistical problems in agricultural experimentation. *J. R. Stat. Soc. Suppl. 2*, 107-154.
48. Ngo, L. and Rand, R. (2002). Model Selection in Linear Mixed Effects Models Using SAS® Proc Mixed. SUGI 22.
49. Ofversten, J. (1993). Exact test for Variance Components in Unbalanced Mixed Linear Models. *Biometrics* 49(1): 45-57.
50. Osborne, B.G., Fox, G.P., Kelly, A.M., and Henry, R.J. (2007). Measurement of Barley Grain Rheology for the Quality Selection of Breeding Material. J. Inst. Brew. 113(2), 135–141, 2007.
51. Patterson, H.D. and Thompson, R. (1971). Recovery of inter–block information when block sizes are unequal. *Biometrika* 58: 545–554.
52. Raudenbush, S. W., and Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.
53. Romney, D., Kaitho, R., Biwott, J., Wambugu, M., Chege, L., Omore, A., Staal S.,Wanjohi, P. and Thorpe, W. (2000). Technology development and field testing: access to credit to allow smallholder dairy farmers in central Kenya to reallocate concentrates during lactation. Paper presented at the 3rd All Africa Conference on Animal Agriculture and 11th Conference of the Egyptian Society of Animal Production, 6-9 November, 2000, Alexandria, Egypt.

55. Reverter,A., Byrne, K.A., and Dalrymple, B.P. (1996). A software program for Bayesian Analysis of Micro-array Gene Expression Data. AAABGVol 15 pp. 90-93.

56. Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects (with discussion). Statistical Science 6: 15 - 51.

57. SAS Institute Inc. (1996). SAS/STATA Software: Changes and Enhancements through Release 6.11-cary, N C: SAS Institute Inc.

58. SAS (1999). Statistical Analysis System user's guide: SAS Institute.

59. SAS. (2000). SAS user's guide, Released 9th ed. Cary, North Carolina: SAS Institute Inc.

60. Serfling, R.J. (1980), Approximation Theorems of Mathematical Statistics, New York: John Wiley & Sons, Inc.

61. Schwartz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics*, 6,461–464.

62. Searle S.R. (1988). Mixed Models and Unbalanced Data: Wherefrom, Whereat, and Whereto? Communications in Statistics-Theory and Methods, 17(4), 935-968.

63. Searle, S. R., Casella, G., and Mcculloch, C.E. (2006). Variance Components. New York: John Wiley & Sons, Inc.

64. Searle, S.R. (1995). An overview of variance components estimation. *Metrica* 42(3-4): 215-230.

65. Self, S.G. and Liang, K.Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions, *Journal of the American Statistical Association,* 82, 605-610.

66. Shaun, R. (2002). An approach to estimating the variance components to unbalanced cluster sampled survey data and simulated data. University of South Africa, A thesis for Master of Science in Statistics.

67. Stephen, M. (2007). Review Methods of Estimating Parameters in Known Linear Mixed - Effective (nlme) Models. University of Nairobi, Kenya, Abstracts of Thesis Completed by ILRI Graduate Fellows – 2007 - MSC.

68. Temesign, Z. (2009/10). Special Topic: Mixed Model. Lecture notes for AAU second year Statistics M.Sc students.

69. USAN (1996). National Agricultural Statistics Service: Agricultural statistics [Online], Available at www.nass.usda.gov/nd/intro74.pdf (verified 19 May 1996).

70. Verbeke, G. and Molenberghs, G. (2000). Linear Mixed Models for Longitudinal Data. Springer-verlage New York, Inc.

71. Yann, C., Andreas, K., David, K., and Teja, T. (2007). Spider diversity in cereal fields: Comparing factors at local, landscape and regional scales. Journal of Biogeography (J. Biogeogr.) (2005) 32, 2007–2014.

72. Yates, F. (1936). A new method of arranging variety trials involving a large number of varieties. *Journal of Agricultural Science* 26: 424–455.

73. Yates, F. (1940). Lattice squares. *Journal of Agricultural Science* 30: 672–687.

74. Wolfinger, R.D., Tobias, R.D., and Sall, J. (1994). Computing Gaussian Likelihood and their Derivatives for General Linear Mixed Model, SIAM Journal on Scientific Computing, 15(6), 1294-1310.

75. Zewotir, T., Galpin, J. S. (2005). Influence diagnostics for linear mixed models. *J. Data Sci.*3:153–177.

76. Zewotir, T. and Galpin, J. S. (2006). Evaluation of Linear Mixed Model Case Deletion Diagnostic Tools by Monte Carlo Simulation. *Communications in Statistics - Simulation and Computation*, 35: 3, 645 —682.

APPENDIX A

Table 1: Variety types in the study

Variety	code
IFBON 1/97	1
IFBON 49/97	2
IFBON 50/97	3
IFBON 118/97	4
IFBON 128/97	5
IWFBSO 74/97 sp	6
IWFBSO 75/97	7
IWFBSO 76/97	8
IWFBSO 86/97	9
IWFBSO 88/97	10
IWFBSO 97/97	11
IWFBSO 132/97	12
BYDV 34/97	13
BYDV 42/97	14
BYDV 45/97	15
BYDV 48/97	16
BYDV 49/97	17
BSP SPS 46/95	18
EH 1607/F6.1H.5.6.2	19
EH 1607/F6.3H.5.8.3	20
EH 1607/F6.4H.5.9.4	21
EH 1607/F6.3H.8.11.5	22
EH 1627/F7.2H.5.102	23
EH 1607/F6.B4.8.13.7	24
EH 1607/F6.3H.22.26.9	25
EH 1622/F6.2H.23.29.10	26
EH 1622/F6 5H.23.33.13	27
EH 1622/F6.6H.23.34.14	28
EH 1644/F7.1H.17.35.15	29
EH 1665/F7.1H.28.40.16	30
EH 1665/F7.2H.28.41.17	31
EH 1665/F7.4H.28.43.19	32
EH 1665/F7.6H.28.46.20	33
EH1668/F7.1H.31.49.22	34
EH 1668/F7.1H.31.51.24	35
EH 1668/F7.6H.29.80	36
EH 1624/F7.1H.2.99	37
EH 1627/F7.1H.5.101	38
EH 1627/F7.2H.5.102	39
EH 1627/F7.4H.5.105	40
EH 1628/F7.1H.6.107	41
EH 1641/F7.1H.13.109	42
EH 1601/F6.B4.21.1	43
EH 1621/F6.B1.22.16.14	44
EH 1622/F6.B1.23.18.16	45
EH 1624/F7.B1.2.20.17	46
EH 1627/F7.B1.5.21.18	47
EH 1628/F7.B1.6.22.19	48
EH 1642/F7.B1.15.25.23	49
EH 1643/F7.B1.16.27.25	50
EH 1654/F7.B1.19.29	51
EH 1655/F7.B1.28.33	52
EH 1666/F7.B5.29.34	53
EH 1668/F7.B1.31.36	54
EH 1671/F7.B2.34.41	55
EH 1689/F7.B1.52.58	56
EH 1699/F7.B2.62.69	57
EH 1700/F7.B1.63.70	58
EH 1702/F7.B2.65.73	59
EH 1772/F7.B1.85.82	60
EH 1737/F7.B3.100.83	61
EH 1748/F7.B3.111.85	62
EH 1676/F7.B1.39.45	63
Local ck.	64

APPENDIX B: PROC MIXED Codes Using SAS.

Code1: /*Influence Diagnostics in the Linear Mixed Model*/
```
ods html;ods graphics on; /*this is used to open the output delivery system*/
proc mixed data=barley IC;
class rep block variety env;
model yield =  env rep(env) block(rep)/ influence(iter=3 keep=50 est);
random  variety  variety*env;run;
ods graphics on;ods html close;
/******************************************************/
```

Code2: /*PROC UNIVARIATE Procedure for Goodness-of-Fit Tests Normal Distribution (Test of Normality).*/

```
ods html;ods graphics on;
proc mixed data= barley method=ml/reml  ;
class rep block variety env;
model yield =  env rep(env) block(rep)/  outp=csfits;
random variety variety*env ;run;
proc univariate data=csfits; /*this procedures used to test normality
including histogram of residual Using ML.*/
var resid;histogram / normal;
title 'Residual Analysis of barley data';run;
ods graphics on;ods html close;
```

Code3: /*PROC UNIVARIATE Procedure for Test of Normality After LN Transformation.*/

```
ods html;ods graphics on;
data logbarley;set barley;logyield=log(yield);run;
proc mixed data= logbarley method=ml/reml  ;
class rep block variety env;
model logyield =  env rep(env) block(rep)/  outp=csfits;
random variety variety*env ;run;
proc univariate data=csfits; /*this procedures used to test normality
including histogram of residual After LN Transformation Using ML.*/
var resid;histogram / normal;
title 'Residual Analysis of logbarley data';run;
ods graphics on;ods html close;

/****************************************************************/
```

Code 4: /*PROC MEAN Procedure using SAS.*/
```
   proc means data = barley; var yield ;run;  /*this procedure prints the mean
                                of over all of yield taht has only one value*/
       proc sort data=barley; by location; proc means data=barley;var yield;by
  location;run; /*this is used to print out the mean of yield in each location
                                                                 with 4 values*/
        proc sort data=barley; by variety; proc means data=barley;var yield;by
     variety;run;/* this is the procedure of calculating the mean of varieties
                                                          that prints 64 values*/
proc sort data=barley; by env; proc means data=barley;var yield;by env;run;
/*here is the procedure that print out 8 values of enviromnents mean*/
/*********************************************************************/
```

Code 5: /*Covariance Parameter Estimate Procedure for SLD and SLD with missing plots.*/

```
     proc varcomp data= barley method =type1; /*Variance Components Estimation
                                     Procedure using Henderson's Method III*/
class rep block variety env;model yield = env rep(env) block(rep)variety
variety*env ;run;

proc mixed data= barley method=ml/reml covtest  asycov ; /*Variance
Components Estimation Procedure using ML*/
class rep block variety env;model yield =  env rep(env) block(rep);
random  variety variety*env/ type =vc;run;

/**********************************************************************/
```

Code 7:
```
/*The GLM procedures taking that variety and variety*environment as
fixed effects for SLD and SLD with missing plots*/
ods html;ods graphics on;
proc glm data= barley;class rep block variety env;
model yield =  env rep(env) block(rep)variety  variety*environment;run;
ods html close;ods graphics on;
/**********************************************************************/
```

Code 8:
```
/*Solution for Fixed Effects, Random Effects and least square means of
fixed effect.*/

proc mixed data= barley method=ml/reml;class rep block variety env;
model yield =  env rep(env) block(rep)/solution;random variety
variety*env/solution;
lsmeans env/diff;run;
/**********************************************************************/
```

APPENDIX C: Partial PROC MIXED Outputs Using SAS.

Constant outputs for all techniques

Model Information	
Data Set	WORK.BARLEY
Dependent Variable	yield
Covariance Structure	Variance Components
Estimation Method	ML/REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Dimensions	
Covariance Parameters	3
Columns in X	41
Columns in Z	576
Subjects	1
Max Obs Per Subject	1024
	Criteria met

Class Level Information		
Class	Levels	Values
rep	2	1 2
Block	8	1 2 3 4 5 6 7 8
Variety	64	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64
env	8	1 2 3 4 5 6 7 8

Summary 1:

Table C1: Top 50 Influence observations for Tuples of size 1 Arranged by Likelihood Distance.

Influence Diagnostics for Tuples of size 1 Arranged by Likelihood Distance												
Deleted Obs.Index	Iter	PRESS Statistic	Cook'sD	MDFFITS	COV-RATIO	COV-TRACE	Cook's D CovParms	MDFFITS CovParms	COVRATIO CovParms	COVTRACE CovParms	RMSEw/o deleted tuple	RLD
140	2	1127.25	0.01784	0.01822	0.8248	0.19	0.26279	0.27606	0.9108	0.0913	6.24083	0.4184
434	2	287.99	0.00772	0.0078	0.9106	0.0925	0.08636	0.08786	1.0029	0.0044	6.28401	0.1503
79	2	729.76	0.00784	0.00783	0.9543	0.0464	0.05205	0.05347	0.9294	0.072	6.26476	0.116
128	2	401.15	0.00737	0.00742	0.9173	0.0857	0.05271	0.05397	0.9784	0.0213	6.28387	0.1133
190	2	812.25	0.00641	0.00645	0.9027	0.1016	0.05244	0.05426	0.9294	0.0716	6.29788	0.1056
229	2	352.8	0.00644	0.0064	1.0049	0.0049	0.05068	0.05205	0.9731	0.027	6.256	0.1032
58	2	330.78	0.00561	0.00556	1.0253	0.0251	0.05636	0.05773	0.9736	0.0263	6.25493	0.1021
558	2	295.67	0.00588	0.00584	1.0185	0.0184	0.05299	0.05439	0.98	0.0199	6.25525	0.101
101	2	382.9	0.0049	0.00495	0.8982	0.1065	0.05956	0.06096	0.983	0.0167	6.31195	0.1006
14	2	344.98	0.00532	0.00527	1.0238	0.0236	0.05101	0.05213	0.9708	0.0292	6.25785	0.0942
270	2	468.49	0.00664	0.00661	0.9787	0.0215	0.03681	0.03758	0.958	0.0425	6.26458	0.0906
55	2	474.51	0.00625	0.00622	0.9821	0.018	0.03425	0.03495	0.9562	0.0443	6.26631	0.0848
130	2	216.48	0.00441	0.00436	1.0389	0.0383	0.04825	0.04913	0.9879	0.0118	6.25937	0.0841
650	2	371.91	0.00467	0.00462	1.0226	0.0225	0.04361	0.04437	0.9662	0.0338	6.26343	0.0814
33	2	204.9	0.00481	0.00477	1.0239	0.0237	0.03966	0.04051	0.9911	0.0087	6.26179	0.0787
596	2	414.04	0.00611	0.0061	0.9672	0.0332	0.02507	0.02543	0.9653	0.0351	6.2733	0.0744
254	2	125.5	0.00356	0.00352	1.0502	0.0491	0.04541	0.04597	1.0016	0.002	6.26198	0.0743
637	2	412.67	0.00371	0.00374	0.9129	0.0906	0.04285	0.04341	0.9769	0.0228	6.31509	0.0735
114	2	69.63	0.00294	0.0029	1.0606	0.0591	0.0468	0.04707	1.0122	0.0126	6.26313	0.0706
121	2	398.59	0.00401	0.00404	0.917	0.0862	0.03714	0.03775	0.9761	0.0238	6.31093	0.0702
645	2	244.79	0.00545	0.00544	0.9807	0.0194	0.0255	0.02583	0.9889	0.0109	6.27318	0.0696

119	2	376.51	0.00348	0.0035	0.9164	0.0868	0.04054	0.04102	0.9809	0.0189	6.31546	0.0692
721	2	376.05	0.0054	0.00542	0.9403	0.0613	0.02226	0.02263	0.973	0.0272	6.28991	0.0661
597	2	81.98	0.0026	0.00256	1.0619	0.0604	0.04416	0.04421	1.0083	0.0087	6.26534	0.0652
82	2	50.99	0.00272	0.00269	1.0579	0.0565	0.04084	0.04103	1.0155	0.0158	6.2658	0.0629
7	2	297.86	0.00513	0.00511	0.9877	0.0123	0.02083	0.02115	0.9779	0.0222	6.27301	0.0622
473	2	36.03	0.00284	0.00285	0.9523	0.0484	0.0386	0.03831	1.0329	0.0335	6.30566	0.0615
738	2	187.05	0.0043	0.00428	1.0118	0.0117	0.02386	0.02427	0.992	0.0078	6.27032	0.0586
70	2	106.7	0.00255	0.00252	1.0536	0.0525	0.03587	0.03595	1.0019	0.0022	6.26873	0.0564
889	2	175.61	0.00457	0.00456	0.9733	0.0269	0.01948	0.01958	0.999	0.0007	6.28322	0.0563
158	2	48.18	0.00201	0.00198	1.063	0.0613	0.03718	0.037	1.0142	0.0145	6.26961	0.0533
63	2	286.99	0.00222	0.00223	0.9309	0.0713	0.03345	0.03344	0.9931	0.0066	6.3194	0.0517
515	2	298.8	0.00215	0.00216	0.9308	0.0714	0.03388	0.03382	0.9925	0.0071	6.31999	0.0515
714	2	23.47	0.0024	0.00238	1.0522	0.0511	0.0321	0.03216	1.0217	0.0219	6.27038	0.0514
233	2	290.68	0.00336	0.00337	0.9358	0.066	0.02338	0.02371	0.9852	0.0147	6.30816	0.0508
30	2	185.56	0.00171	0.00172	0.9345	0.0674	0.03575	0.0355	1.009	0.0094	6.32193	0.0498
276	2	324.61	0.00453	0.00452	0.9814	0.0187	0.01313	0.01327	0.9729	0.0274	6.28021	0.0496
784	2	110.27	0.00385	0.00384	0.9812	0.0188	0.01754	0.0175	1.0084	0.0089	6.28581	0.0484
568	2	402.33	0.00334	0.00335	0.9383	0.0634	0.02048	0.02075	0.9712	0.0289	6.30728	0.0476
47	2	185.1	0.00403	0.00403	0.9682	0.0322	0.01363	0.01374	0.9953	0.0045	6.28952	0.0461
719	2	465.15	0.00371	0.00371	0.951	0.05	0.01504	0.01534	0.9608	0.0395	6.29904	0.0451
539	2	147.8	0.00276	0.00273	1.0333	0.0329	0.02124	0.02142	0.993	0.0069	6.27446	0.0434
212	2	171.77	0.00393	0.00392	0.9811	0.019	0.01181	0.01188	0.9954	0.0044	6.28518	0.0434
667	2	286.5	0.00408	0.00408	0.9658	0.0347	0.00965	0.00973	0.9791	0.021	6.29005	0.0425
644	2	276.37	0.00397	0.00395	0.9843	0.0158	0.00924	0.00932	0.978	0.0221	6.28361	0.0411
640	2	84.55	0.00261	0.00259	1.0353	0.0348	0.01998	0.02015	1.0039	0.004	6.27493	0.041
435	2	327.9	0.00313	0.00314	0.9461	0.0552	0.01556	0.01574	0.9779	0.0222	6.30571	0.0409
634	2	232.36	0.00252	0.00253	0.9441	0.0573	0.02011	0.02027	0.992	0.0079	6.31144	0.0406
2	2	165.36	0.00253	0.00253	0.9491	0.052	0.01893	0.01911	1.0001	0.0002	6.30933	0.0395
249	2	229.96	0.00307	0.00305	1.0141	0.0141	0.01379	0.01394	0.9812	0.0188	6.27919	0.0385

Summary 2a: SAS Procedure Output for Test of Normality Using ML and REML.

Moments			
N	1024	Sum Weights	1024
Mean	0	Sum Observations	0
Std Deviation	4.97080076	Variance	24.7088602
Skewness	**0.09005799**	Kurtosis	0.54183697
Uncorrected SS	25277.164	Corrected SS	25277.164
Coeff Variation	.	Std Error Mean	0.15533752

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	-14	Pr >= |M|	0.3988
Signed Rank	S	-3556	Pr >= |S|	0.7074

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.02305532	Pr > D	>0.150
Cramer-von Mises	W-Sq	0.08949345	Pr > W-Sq	**0.159**
Anderson-Darling	A-Sq	0.59091074	Pr > A-Sq	0.128

Summary 2b: SAS Procedure Output for Test of Normality after LN Transformation Using ML and REML.

Moments			
N	1024	Sum Weights	1024
Mean	0	Sum Observations	0
Std Deviation	0.27044114	Variance	0.07313841
Skewness	**-0.7695897**	Kurtosis	2.1614283
Uncorrected SS	74.8205911	Corrected SS	74.8205911
Coeff Variation	.	Std Error Mean	0.00845129

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	0	Pr > |t|	1.0000
Sign	M	32	Pr >= |M|	0.0489
Signed Rank	S	16680	Pr >= |S|	0.0781

Goodness-of-Fit Tests for Normal Distribution				
Test	Statistic		p Value	
Kolmogorov-Smirnov	D	0.06849124	Pr > D	<0.010
Cramer-von Mises	W-Sq	1.15266017	Pr > W-Sq	**<0.005**
Anderson-Darling	A-Sq	7.45081855	Pr > A-Sq	<0.005

Summary 3: Descriptive Statistics for all variables in the study.

variables	N	Mean	Std Dev	Minimum	Maximum
overall	1024	25.2703906	13.7141413	1.2000000	70.2500000
location=1	256	43.8112891	8.9343567	10.9000000	70.2500000
location=2	256	13.7828125	6.9330437	1.2000000	34.8000000
location=3	256	23.2781250	9.0458391	3.9000000	45.3000000
location=4	256	20.2093359	6.1246998	4.7500000	39.7500000
variety=1	16	22.9875000	17.0416981	2.0000000	56.0400000
variety=2	16	19.4418750	13.7352317	2.2000000	42.3000000
variety=3	16	19.2618750	13.6805878	4.2000000	50.6000000
variety=4	16	23.0287500	15.2824631	2.8500000	51.0100000
variety=5	16	23.4087500	12.8778227	8.9000000	50.1000000
variety=6	16	19.9837500	15.2921108	5.1500000	45.8900000
variety=7	16	23.5981250	19.4986725	2.0000000	63.2700000
variety=8	16	26.2993750	19.5440090	2.1500000	62.8400000
variety=9	16	21.0987500	17.1201557	3.6000000	51.0600000
variety=10	16	21.0537500	15.6038508	2.3500000	46.9300000
variety=11	16	22.7931250	18.0656948	2.1000000	56.8900000
variety=12	16	17.0643750	11.1058429	1.2000000	43.1200000
variety=13	16	22.4300000	15.0107179	1.9000000	50.0900000
variety=14	16	18.9143750	9.3527229	5.1500000	34.7500000
variety=15	16	20.7250000	11.8089983	3.4000000	44.0000000
variety=16	16	14.7893750	8.3114319	3.9000000	32.0800000
variety=17	16	19.6681250	10.1204441	6.2000000	40.3000000
variety=18	16	24.2481250	16.9050163	2.2500000	56.1700000
variety=19	16	24.4106250	11.1093963	7.1500000	46.4600000
variety=20	16	24.0643750	9.8203170	10.9000000	42.1800000
variety=21	16	27.7543750	15.6089530	9.3500000	53.2400000
variety=22	16	25.0175000	12.3024092	8.6000000	48.5800000
variety=23	16	30.2981250	12.6411045	11.6000000	51.3600000
variety=24	16	28.2793750	15.0987277	10.8000000	52.9100000
variety=25	16	24.8968750	10.3651613	10.1000000	40.1700000
variety=26	16	26.2506250	11.8317778	6.9000000	53.6400000
variety=27	16	26.7062500	12.5549241	10.5000000	50.1000000
variety=28	16	25.9212500	14.0478766	5.7000000	52.5300000
variety=29	16	30.2887500	12.5168904	13.1500000	60.6300000
variety=30	16	24.2981250	13.3799239	7.8000000	49.1900000
variety=31	16	22.4518750	14.0890360	2.1000000	55.1500000
variety=32	16	24.2643750	12.1884995	10.6000000	52.2100000
variety=33	16	27.9400000	11.5931267	10.6000000	47.2000000
variety=34	16	27.6606250	12.7618455	9.0500000	53.7300000
variety=35	16	28.0118750	13.6137397	9.0000000	54.2500000
variety=36	16	27.8131250	13.6492128	11.8500000	55.3600000
variety=37	16	31.1868750	14.4595720	8.9500000	55.9600000
variety=38	16	28.8400000	12.6450053	9.5500000	61.8400000
variety=39	16	28.0731250	11.6039331	11.0500000	50.9600000
variety=40	16	21.7450000	12.0955953	6.9500000	43.4300000
variety=41	16	21.9975000	8.9518084	6.3000000	36.5100000
variety=42	16	29.0943750	17.3873391	4.7500000	63.2500000
variety=43	16	26.7712500	13.8376125	7.8000000	51.9600000

variety=44	16	25.9693750	12.9702647	4.6500000	51.6000000
variety=45	16	25.7056250	13.0830399	6.9500000	46.7000000
variety=46	16	29.6500000	12.7579110	4.4500000	49.4600000
variety=47	16	27.7906250	14.7518869	4.9500000	61.4500000
variety=48	16	21.2012500	7.0442675	9.1000000	34.3400000
variety=49	16	29.9393750	12.6692517	11.7000000	50.3500000
variety=50	16	31.9381250	14.4244497	10.4500000	61.8700000
variety=51	16	24.5443750	10.2612325	14.6000000	43.0600000
variety=52	16	27.1362500	14.0390284	7.3000000	48.4600000
variety=53	16	25.2343750	11.6721006	9.5000000	50.2500000
variety=54	16	28.5787500	13.0328185	6.0000000	53.6900000
variety=55	16	27.3212500	16.4702301	6.7000000	61.1500000
variety=56	16	28.2818750	15.0461832	6.2000000	55.1400000
variety=57	16	22.8537500	11.5196174	8.6000000	43.1400000
variety=58	16	33.2056250	17.8065040	7.7000000	70.2500000
variety=59	16	27.4025000	13.3646661	8.3500000	52.0100000
variety=60	16	25.8431250	12.8066018	9.3500000	49.2400000
variety=50	16	24.4556250	10.3732425	7.8000000	40.9400000
variety=50	16	31.4275000	17.0222922	7.2000000	58.9100000
variety=50	16	24.4856250	11.8729131	11.9500000	45.8400000
Variety=64	16	29.5087500	12.0811643	11.3000000	57.7900000
env=1	128	43.4088281	9.0449658	21.7800000	70.2500000
env=2	128	44.2137500	8.8394614	10.9000000	62.8400000
env=3	128	10.7644531	6.1638205	1.2000000	32.2500000
env=4	128	16.8011719	6.3368391	3.6000000	34.8000000
env=5	128	22.7046875	9.0394116	3.9000000	42.9000000
env=6	128	23.8515625	9.0512385	4.7000000	45.3000000
env=7	128	17.8699219	5.4572907	4.7500000	28.3000000
env=8	128	22.5487500	5.8741248	9.5500000	39.7500000

Summary 4: The GLM Procedure taking to that variety and variety*environment as fixed effects.

Type I and Type II Sum of Squares with Root MSE and Coefficient of Variation for SLD Data.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	533	175980.0822	330.1690	9.85	<.0001
Error	490	16423.3746	33.5171		
Corrected Total	1023	192403.4568			
Source	**DF**	**Type I SS**	**Mean Square**	**F Value**	**Pr > F**
env	7	133219.0515	19031.2931	567.81	<.0001
rep(env)	8	3150.7198	393.8400	11.75	<.0001
Block(rep)	14	1114.5813	79.6129	2.38	0.0033
Variety	63	14157.3840	224.7204	6.70	<.0001
Variety*env	441	24338.3456	55.1890	1.65	<.0001

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
env	7	133219.0515	19031.2931	567.81	<.0001
rep(env)	7	2564.4747	366.3535	10.93	<.0001
Block(rep)	14	750.2103	53.5865	1.60	0.0755
Variety	63	14157.3840	224.7204	6.70	<.0001
Variety*env	441	24338.3456	55.1890	1.65	<.0001
Root MSE		5.789395	Coeff var		22.9098

Type I and Type II Sum of Squares with Root MSE and Coefficient of Variation for PBLD Data.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	521	169501.9903	325.3397	9.73	<.0001
Error	474	15842.5230	33.4230		
Corrected Total	995	185344.5133			
Source	**DF**	**Type I SS**	**Mean Square**	**F Value**	**Pr > F**
env	7	127761.1367	18251.5910	546.08	<.0001
rep(env)	8	3048.6213	381.0777	11.40	<.0001
Block(rep)	14	1078.7902	77.0564	2.31	0.0045
Variety	63	14095.8753	223.7441	6.69	<.0001
Variety*env	429	23517.5668	54.8195	1.64	<.0001

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Source	**DF**	**Type III SS**	**Mean Square**	**F Value**	**Pr > F**
env	7	126242.5183	18034.6455	539.59	<.0001
rep(env)	7	2450.8592	350.1227	10.48	<.0001
Block(rep)	14	634.2197	45.3014	1.36	0.1712
Variety	63	14031.0568	222.7152	6.66	<.0001
Variety*env	429	23517.5668	54.8195	1.64	<.0001
Root MSE		5.781267	Coeff Var		22.89121

Summary 5: The SAS output for Predictions of Random Effects and differences of Least Squares Means of fixed effects Using ML and REML.

Table C2: Differences of Least Squares Means for Fixed Effects.

Differences of Least Squares Means							
Effect	env	_env	Estimate	Standard Error	DF	t Value	Pr > |t|
env	1	2	-0.8049	0.7729	764	-1.04	0.2980
env	1	3	32.6444	0.9598	764	34.01	<.0001
env	1	4	26.6077	0.9598	764	27.72	<.0001
env	1	5	20.7041	0.9598	764	21.57	<.0001
env	1	6	19.5573	0.9598	764	20.38	<.0001
env	1	7	25.5389	0.9598	764	26.61	<.0001
env	1	8	20.8601	0.9598	764	21.73	<.0001
env	2	3	33.4493	0.9598	764	34.85	<.0001
env	2	4	27.4126	0.9598	764	28.56	<.0001
env	2	5	21.5091	0.9598	764	22.41	<.0001
env	2	6	20.3622	0.9598	764	21.21	<.0001
env	2	7	26.3438	0.9598	764	27.45	<.0001
env	2	8	21.6650	0.9598	764	22.57	<.0001
env	3	4	-6.0367	0.7729	764	-7.81	<.0001

Differences of Least Squares Means							
Effect	env	_env	Estimate	Standard Error	DF	t Value	Pr > |t|
env	3	5	-11.9402	0.9598	764	-12.44	<.0001
env	3	6	-13.0871	0.9598	764	-13.63	<.0001
env	3	7	-7.1055	0.9598	764	-7.40	<.0001
env	3	8	-11.7843	0.9598	764	-12.28	<.0001
env	4	5	-5.9035	0.9598	764	-6.15	<.0001
env	4	6	-7.0504	0.9598	764	-7.35	<.0001
env	4	7	-1.0688	0.9598	764	-1.11	0.2659
env	4	8	-5.7476	0.9598	764	-5.99	<.0001
env	5	6	-1.1469	0.7729	764	-1.48	0.1383
env	5	7	4.8348	0.9598	764	5.04	<.0001
env	5	8	0.1559	0.9598	764	0.16	0.8710
env	6	7	5.9816	0.9598	764	6.23	<.0001
env	6	8	1.3028	0.9598	764	1.36	0.1751
env	7	8	-4.6788	0.7729	764	-6.05	<.0001

Table C2: Empirical Best Unbiased Linear Estimations for rep(env) and block(rep).

Solution for Fixed Effects (rep(env) and block(rep))								
Effect	rep	Block	env	Estimate	Standard Error	DF	t Value	Pr > |t|
rep(env)	1		1	-2.4536	1.6389	490	-1.5	0.135
rep(env)	2		1	0	.	.	.	.
rep(env)	1		2	-1.9528	1.6389	490	-1.19	0.234
rep(env)	2		2	0	.	.	.	.
rep(env)	1		3	2.2602	1.6389	490	1.38	0.1685
rep(env)	2		3	0	.	.	.	.
rep(env)	1		4	-2.6429	1.6389	490	-1.61	0.1075
rep(env)	2		4	0	.	.	.	.
rep(env)	1		5	-4.8593	1.6389	490	-2.96	0.0032
rep(env)	2		5	0	.	.	.	.
rep(env)	1		6	-6.9046	1.6389	490	-4.21	<.0001
rep(env)	2		6	0	.	.	.	.
rep(env)	1		7	3.4071	1.6389	490	2.08	0.0382
rep(env)	2		7	0	.	.	.	.
rep(env)	1		8	-1.7975	1.6389	490	-1.1	0.2733
rep(env)	2		8	0	.	.	.	.
B(rep)	1	1		2.0175	1.4237	490	1.42	0.1571
B(rep)	1	2		-0.1982	1.3523	490	-0.15	0.8836
B(rep)	1	3		1.635	1.3474	490	1.21	0.2255
B(rep)	1	4		-0.1764	1.3616	490	-0.13	0.897
B(rep)	1	5		1.425	1.3607	490	1.05	0.2955
B(rep)	1	6		1.0789	1.4101	490	0.77	0.4446
B(rep)	1	7		3.1395	1.3466	490	2.33	0.0201
B(rep)	1	8		0	.	.	.	.
B(rep)	2	1		2.1483	1.3652	490	1.57	0.1162

B(rep)	2	2		1.5725	1.4012	490	1.12	0.2623
B(rep)	2	3		-0.6455	1.4299	490	-0.45	0.6519
B(rep)	2	4		1.0807	1.3956	490	0.77	0.4391
B(rep)	2	5		1.8671	1.3884	490	1.34	0.1793
B(rep)	2	6		0.854	1.3768	490	0.62	0.5354
B(rep)	2	7		-0.7927	1.4088	490	-0.56	0.5739
B(rep)	2	8		0	.	.	.	.

Table C3: Empirical Best Unbiased Linear Predictions for varieties.

Effect	variety	Estimate	Std Err Pred	DF	t Vale	Pr > |t|
variety	58	6.0353	1.6464	512	3.67	0.0003
variety	50	5.0713	1.6464	512	3.08	0.0022
variety	62	4.6829	1.6464	512	2.84	0.0046
variety	37	4.4999	1.6464	512	2.73	0.0065
variety	23	3.824	1.6464	512	2.32	0.0206
variety	29	3.8168	1.6464	512	2.32	0.0208
variety	49	3.5511	1.6464	512	2.16	0.0315
variety	46	3.331	1.6464	512	2.02	0.0436
variety	64	3.2236	1.6464	512	1.96	0.0508
variety	42	2.9084	1.6464	512	1.77	0.0779
variety	38	2.7149	1.6464	512	1.65	0.0998
variety	54	2.5162	1.6464	512	1.53	0.127
variety	56	2.2905	1.6464	512	1.39	0.1648
variety	24	2.2886	1.6464	512	1.39	0.1651
variety	39	2.1317	1.6464	512	1.29	0.196
variety	35	2.0851	1.6464	512	1.27	0.2059
variety	33	2.0304	1.6464	512	1.23	0.218
variety	36	1.9339	1.6464	512	1.17	0.2407
variety	47	1.9168	1.6464	512	1.16	0.2449
variety	21	1.8893	1.6464	512	1.15	0.2517
variety	34	1.8179	1.6464	512	1.1	0.27
variety	59	1.6216	1.6464	512	0.98	0.3251
variety	55	1.5598	1.6464	512	0.95	0.3439
variety	52	1.4191	1.6464	512	0.86	0.3891
variety	43	1.1415	1.6464	512	0.69	0.4884
variety	27	1.0921	1.6464	512	0.66	0.5074
variety	8	0.7826	1.6464	512	0.48	0.6347
variety	26	0.7455	1.6464	512	0.45	0.6509
variety	44	0.5316	1.6464	512	0.32	0.7469
variety	28	0.495	1.6464	512	0.3	0.7638
variety	60	0.4356	1.6464	512	0.26	0.7914
variety	45	0.331	1.6464	512	0.2	0.8407
variety	53	-0.0274	1.6464	512	-0	0.9867

variety	22	-0.1923	1.6464	512	-0.1	0.907
variety	25	-0.2841	1.6464	512	-0.2	0.8631
variety	51	-0.5522	1.6464	512	-0.3	0.7375
variety	63	-0.5969	1.6464	512	-0.4	0.7171
variety	61	-0.6197	1.6464	512	-0.4	0.7068
variety	19	-0.6539	1.6464	512	-0.4	0.6914
variety	30	-0.7395	1.6464	512	-0.5	0.6535
variety	32	-0.7651	1.6464	512	-0.5	0.6423
variety	18	-0.7775	1.6464	512	-0.5	0.637
variety	20	-0.9173	1.6464	512	-0.6	0.5777
variety	7	-1.2719	1.6464	512	-0.8	0.4402
variety	5	-1.4159	1.6464	512	-0.9	0.3902
variety	4	-1.7049	1.6464	512	-1	0.3009
variety	1	-1.7363	1.6464	512	-1.1	0.2921
variety	57	-1.838	1.6464	512	-1.1	0.2648
variety	11	-1.8841	1.6464	512	-1.1	0.253
variety	31	-2.1437	1.6464	512	-1.3	0.1935
variety	13	-2.1603	1.6464	512	-1.3	0.1901
variety	41	-2.4893	1.6464	512	-1.5	0.1312
variety	40	-2.6813	1.6464	512	-1.6	0.104
variety	48	-3.0949	1.6464	512	-1.9	0.0607
variety	9	-3.1728	1.6464	512	-1.9	0.0545
variety	10	-3.2071	1.6464	512	-2	0.052
variety	15	-3.4571	1.6464	512	-2.1	0.0362
variety	6	-4.0209	1.6464	512	-2.4	0.0149
variety	17	-4.2609	1.6464	512	-2.6	0.0099
variety	2	-4.433	1.6464	512	-2.7	0.0073
variety	3	-4.5699	1.6464	512	-2.8	0.0057
variety	14	-4.8342	1.6464	512	-2.9	0.0035
variety	12	-6.2413	1.6464	512	-3.8	0.0002
variety	16	-7.9716	1.6464	512	-4.8	<.0001

Table C4: Empirical Best Unbiased Linear Predictions for variety*environment.

Effect	var	env	Estimate	Std Err Pred	DF	t Value	Pr > |t|
var*env	8	1	4.2291	2.3581	512	1.79	0.0735
var*env	42	1	3.864	2.3581	512	1.64	0.1019
var*env	7	1	3.4251	2.3581	512	1.45	0.147
var*env	18	1	3.1546	2.3581	512	1.34	0.1816
var*env	58	1	2.5695	2.3581	512	1.09	0.2764
var*env	11	1	2.2663	2.3581	512	0.96	0.337
var*env	1	1	2.1713	2.3581	512	0.92	0.3576
var*env	9	1	1.8836	2.3581	512	0.8	0.4248
var*env	4	1	1.7333	2.3581	512	0.74	0.4626
var*env	21	1	1.7295	2.3581	512	0.73	0.4636
var*env	35	1	1.5609	2.3581	512	0.66	0.5083
var*env	6	1	1.3121	2.3581	512	0.56	0.5782
var*env	24	1	1.2315	2.3581	512	0.52	0.6017
var*env	10	1	1.1576	2.3581	512	0.49	0.6237
var*env	62	1	1.0943	2.3581	512	0.46	0.6428
var*env	56	1	1.0101	2.3581	512	0.43	0.6686

var*env	52	1	0.9378	2.3581	512	0.4	0.691
var*env	47	1	0.8037	2.3581	512	0.34	0.7334
var*env	38	1	0.8029	2.3581	512	0.34	0.7336
var*env	59	1	0.7338	2.3581	512	0.31	0.7558
var*env	22	1	0.6985	2.3581	512	0.3	0.7672
var*env	64	1	0.5945	2.3581	512	0.25	0.801
var*env	63	1	0.565	2.3581	512	0.24	0.8107
var*env	44	1	0.5035	2.3581	512	0.21	0.831
var*env	36	1	0.3292	2.3581	512	0.14	0.889
var*env	37	1	0.3033	2.3581	512	0.13	0.8977
var*env	2	1	0.2166	2.3581	512	0.09	0.9269
var*env	13	1	0.1202	2.3581	512	0.05	0.9594
var*env	32	1	0.02729	2.3581	512	0.01	0.9908
var*env	34	1	0.00626	2.3581	512	0	0.9979
var*env	12	1	-0.0034	2.3581	512	0	0.9989
var*env	40	1	-0.0276	2.3581	512	-0	0.9907
var*env	50	1	-0.0485	2.3581	512	-0	0.9836
var*env	30	1	-0.1628	2.3581	512	-0.1	0.945
var*env	43	1	-0.1778	2.3581	512	-0.1	0.9399
var*env	57	1	-0.2628	2.3581	512	-0.1	0.9113
var*env	31	1	-0.47	2.3581	512	-0.2	0.8421
var*env	46	1	-0.6584	2.3581	512	-0.3	0.7802
var*env	28	1	-0.6878	2.3581	512	-0.3	0.7707
var*env	26	1	-0.7324	2.3581	512	-0.3	0.7562
var*env	54	1	-0.7622	2.3581	512	-0.3	0.7467
var*env	27	1	-0.8593	2.3581	512	-0.4	0.7157
var*env	5	1	-0.9057	2.3581	512	-0.4	0.7011
var*env	23	1	-0.942	2.3581	512	-0.4	0.6897
var*env	17	1	-1.0551	2.3581	512	-0.5	0.6547
var*env	53	1	-1.0682	2.3581	512	-0.5	0.6507

var*env	25	1	-1.0947	2.3581	512	-0.5	0.6427
var*env	20	1	-1.1208	2.3581	512	-0.5	0.6348
var*env	45	1	-1.1432	2.3581	512	-0.5	0.628
var*env	60	1	-1.2723	2.3581	512	-0.5	0.5897
var*env	39	1	-1.2996	2.3581	512	-0.6	0.5818
var*env	3	1	-1.3867	2.3581	512	-0.6	0.5568
var*env	55	1	-1.4768	2.3581	512	-0.6	0.5314
var*env	33	1	-1.5968	2.3581	512	-0.7	0.4986
var*env	51	1	-1.6284	2.3581	512	-0.7	0.4901
var*env	19	1	-1.7182	2.3581	512	-0.7	0.4666
var*env	29	1	-1.7574	2.3581	512	-0.8	0.4564
var*env	15	1	-1.9018	2.3581	512	-0.8	0.4203
var*env	16	1	-1.9423	2.3581	512	-0.8	0.4105
var*env	61	1	-2.0277	2.3581	512	-0.9	0.3902
var*env	41	1	-2.0803	2.3581	512	-0.9	0.3781
var*env	14	1	-2.118	2.3581	512	-0.9	0.3695
var*env	49	1	-2.6136	2.3581	512	-1.1	0.2682
var*env	48	1	-4.0326	2.3581	512	-1.7	0.0878
var*env	55	2	3.4031	2.3581	512	1.44	0.1496
var*env	11	2	2.8346	2.3581	512	1.2	0.2299
var*env	7	2	2.158	2.3581	512	0.92	0.3605
var*env	58	2	1.942	2.3581	512	0.82	0.4106
var*env	9	2	1.9172	2.3581	512	0.81	0.4166
var*env	62	2	1.816	2.3581	512	0.77	0.4416
var*env	29	2	1.8056	2.3581	512	0.77	0.4442
var*env	43	2	1.4946	2.3581	512	0.63	0.5265
var*env	50	2	1.4327	2.3581	512	0.61	0.5437
var*env	3	2	1.39	2.3581	512	0.59	0.5558
var*env	8	2	1.3745	2.3581	512	0.58	0.5602
var*env	10	2	1.3596	2.3581	512	0.58	0.5645

var*env	37	2	1.359	2.3581	512	0.58	0.5646
var*env	56	2	1.3198	2.3581	512	0.56	0.5759
var*env	13	2	1.3106	2.3581	512	0.56	0.5786
var*env	31	2	1.1782	2.3581	512	0.5	0.6175
var*env	36	2	0.9674	2.3581	512	0.41	0.6818
var*env	6	2	0.9068	2.3581	512	0.38	0.7007
var*env	30	2	0.8876	2.3581	512	0.38	0.7068
var*env	21	2	0.8852	2.3581	512	0.38	0.7075
var*env	5	2	0.826	2.3581	512	0.35	0.7263
var*env	42	2	0.7723	2.3581	512	0.33	0.7434
var*env	60	2	0.6869	2.3581	512	0.29	0.7709
var*env	24	2	0.5757	2.3581	512	0.24	0.8072
var*env	23	2	0.5689	2.3581	512	0.24	0.8095
var*env	39	2	0.4644	2.3581	512	0.2	0.8439
var*env	59	2	0.4119	2.3581	512	0.17	0.8614
var*env	45	2	0.4054	2.3581	512	0.17	0.8636
var*env	49	2	0.3757	2.3581	512	0.16	0.8735
var*env	28	2	0.3262	2.3581	512	0.14	0.89
var*env	46	2	0.3259	2.3581	512	0.14	0.8901
var*env	4	2	0.2979	2.3581	512	0.13	0.8995
var*env	54	2	0.08214	2.3581	512	0.03	0.9722
var*env	53	2	0.07369	2.3581	512	0.03	0.9751
var*env	47	2	-0.2331	2.3581	512	-0.1	0.9213
var*env	51	2	-0.2468	2.3581	512	-0.1	0.9167
var*env	18	2	-0.2616	2.3581	512	-0.1	0.9117
var*env	35	2	-0.2677	2.3581	512	-0.1	0.9097
var*env	33	2	-0.366	2.3581	512	-0.2	0.8767
var*env	44	2	-0.3973	2.3581	512	-0.2	0.8663
var*env	27	2	-0.4486	2.3581	512	-0.2	0.8492
var*env	63	2	-0.4758	2.3581	512	-0.2	0.8402
var*env	57	2	-0.5159	2.3581	512	-0.2	0.8269
var*env	40	2	-0.5245	2.3581	512	-0.2	0.8241
var*env	19	2	-0.6194	2.3581	512	-0.3	0.7929
var*env	52	2	-0.7064	2.3581	512	-0.3	0.7646
var*env	15	2	-0.9552	2.3581	512	-0.4	0.6856
var*env	61	2	-1.0258	2.3581	512	-0.4	0.6637
var*env	22	2	-1.1625	2.3581	512	-0.5	0.6222
var*env	25	2	-1.236	2.3581	512	-0.5	0.6004
var*env	32	2	-1.4351	2.3581	512	-0.6	0.5431
var*env	26	2	-1.4447	2.3581	512	-0.6	0.5404
var*env	34	2	-1.8776	2.3581	512	-0.8	0.4263
var*env	64	2	-2.0689	2.3581	512	-0.9	0.3807
var*env	48	2	-2.1407	2.3581	512	-0.9	0.3644
var*env	41	2	-2.2055	2.3581	512	-0.9	0.3501
var*env	12	2	-2.3868	2.3581	512	-1	0.3119
var*env	38	2	-2.4544	2.3581	512	-1	0.2984
var*env	17	2	-2.5148	2.3581	512	-1.1	0.2867
var*env	14	2	-2.6795	2.3581	512	-1.1	0.2564
var*env	16	2	-3.8652	2.3581	512	-1.6	0.1018
var*env	20	2	-4.8372	2.3581	512	-2.1	0.0407
var*env	63	2	-0.4758	2.3581	512	-0.2	0.8402
var*env	64	2	-2.0689	2.3581	512	-0.9	0.3807
var*env	20	3	2.8218	2.3581	512	1.2	0.232
var*env	51	3	2.4474	2.3581	512	1.04	0.2998
var*env	21	3	2.2208	2.3581	512	0.94	0.3467
var*env	38	3	1.9984	2.3581	512	0.85	0.3971
var*env	39	3	1.8728	2.3581	512	0.79	0.4275
var*env	36	3	1.751	2.3581	512	0.74	0.4581
var*env	5	3	1.7105	2.3581	512	0.73	0.4685
var*env	34	3	1.6071	2.3581	512	0.68	0.4958

var*env	29	3	1.5872	2.3581	512	0.67	0.5012
var*env	59	3	1.3772	2.3581	512	0.58	0.5594
var*env	49	3	1.2279	2.3581	512	0.52	0.6028
var*env	22	3	1.0174	2.3581	512	0.43	0.6663
var*env	61	3	0.9507	2.3581	512	0.4	0.687
var*env	17	3	0.8406	2.3581	512	0.36	0.7216
var*env	40	3	0.7855	2.3581	512	0.33	0.7392
var*env	32	3	0.7475	2.3581	512	0.32	0.7514
var*env	63	3	0.7426	2.3581	512	0.31	0.7529
var*env	42	3	0.7277	2.3581	512	0.31	0.7577
var*env	48	3	0.6882	2.3581	512	0.29	0.7705
var*env	16	3	0.6146	2.3581	512	0.26	0.7945
var*env	23	3	0.5889	2.3581	512	0.25	0.8029
var*env	3	3	0.4122	2.3581	512	0.17	0.8613
var*env	52	3	0.3007	2.3581	512	0.13	0.8986
var*env	57	3	0.2959	2.3581	512	0.13	0.9002
var*env	18	3	0.1988	2.3581	512	0.08	0.9329
var*env	50	3	0.125	2.3581	512	0.05	0.9577
var*env	53	3	0.09101	2.3581	512	0.04	0.9692
var*env	35	3	0.08765	2.3581	512	0.04	0.9704
var*env	25	3	0.02549	2.3581	512	0.01	0.9914
var*env	6	3	-0.1195	2.3581	512	-0.1	0.9596
var*env	31	3	-0.1268	2.3581	512	-0.1	0.9571
var*env	30	3	-0.1414	2.3581	512	-0.1	0.9522
var*env	43	3	-0.1565	2.3581	512	-0.1	0.9471
var*env	2	3	-0.1633	2.3581	512	-0.1	0.9448
var*env	14	3	-0.1697	2.3581	512	-0.1	0.9426
var*env	27	3	-0.2239	2.3581	512	-0.1	0.9244
var*env	44	3	-0.3221	2.3581	512	-0.1	0.8914
var*env	33	3	-0.3487	2.3581	512	-0.2	0.8825

var*env	12	3	-0.3765	2.3581	512	-0.2	0.8732
var*env	41	3	-0.3771	2.3581	512	-0.2	0.873
var*env	15	3	-0.379	2.3581	512	-0.2	0.8724
var*env	64	3	-0.3805	2.3581	512	-0.2	0.8719
var*env	60	3	-0.404	2.3581	512	-0.2	0.864
var*env	19	3	-0.4068	2.3581	512	-0.2	0.8631
var*env	24	3	-0.5933	2.3581	512	-0.3	0.8014
var*env	26	3	-0.6221	2.3581	512	-0.3	0.792
var*env	1	3	-0.6539	2.3581	512	-0.3	0.7816
var*env	45	3	-0.6855	2.3581	512	-0.3	0.7714
var*env	9	3	-0.8461	2.3581	512	-0.4	0.7199
var*env	56	3	-1.0045	2.3581	512	-0.4	0.6703
var*env	4	3	-1.0663	2.3581	512	-0.5	0.6513
var*env	28	3	-1.0731	2.3581	512	-0.5	0.6493
var*env	54	3	-1.1192	2.3581	512	-0.5	0.6353
var*env	10	3	-1.2004	2.3581	512	-0.5	0.6109
var*env	37	3	-1.2898	2.3581	512	-0.6	0.5846
var*env	55	3	-1.5012	2.3581	512	-0.6	0.5246
var*env	11	3	-1.5028	2.3581	512	-0.6	0.5242
var*env	13	3	-1.5564	2.3581	512	-0.7	0.5095
var*env	46	3	-1.5877	2.3581	512	-0.7	0.5011
var*env	58	3	-1.636	2.3581	512	-0.7	0.4881
var*env	8	3	-1.7363	2.3581	512	-0.7	0.4619
var*env	7	3	-1.8832	2.3581	512	-0.8	0.4249
var*env	47	3	-2.0754	2.3581	512	-0.9	0.3792
var*env	62	3	-2.1336	2.3581	512	-0.9	0.366
var*env	14	4	3.5503	2.3581	512	1.51	0.1328
var*env	46	4	3.1019	2.3581	512	1.32	0.189
var*env	16	4	2.9342	2.3581	512	1.24	0.2139
var*env	57	4	2.326	2.3581	512	0.99	0.3244

var*env	64	4	2.269	2.3581	512	0.96	0.3364
var*env	30	4	2.0975	2.3581	512	0.89	0.3742
var*env	12	4	2.0576	2.3581	512	0.87	0.3833
var*env	35	4	1.8081	2.3581	512	0.77	0.4436
var*env	43	4	1.5976	2.3581	512	0.68	0.4984
var*env	29	4	1.4158	2.3581	512	0.6	0.5485
var*env	48	4	1.4122	2.3581	512	0.6	0.5495
var*env	63	4	1.1636	2.3581	512	0.49	0.6219
var*env	54	4	1.1466	2.3581	512	0.49	0.627
var*env	53	4	1.1112	2.3581	512	0.47	0.6377
var*env	41	4	0.8182	2.3581	512	0.35	0.7288
var*env	31	4	0.7319	2.3581	512	0.31	0.7564
var*env	10	4	0.7153	2.3581	512	0.3	0.7618
var*env	19	4	0.6875	2.3581	512	0.29	0.7708
var*env	28	4	0.6676	2.3581	512	0.28	0.7772
var*env	5	4	0.5628	2.3581	512	0.24	0.8115
var*env	26	4	0.5597	2.3581	512	0.24	0.8125
var*env	33	4	0.4695	2.3581	512	0.2	0.8423
var*env	15	4	0.4662	2.3581	512	0.2	0.8434
var*env	37	4	0.3633	2.3581	512	0.15	0.8776
var*env	51	4	0.2696	2.3581	512	0.11	0.909
var*env	17	4	0.1373	2.3581	512	0.06	0.9536
var*env	50	4	0.1017	2.3581	512	0.04	0.9656
var*env	2	4	0.0961	2.3581	512	0.04	0.9675
var*env	56	4	0.04265	2.3581	512	0.02	0.9856
var*env	44	4	-0.0223	2.3581	512	-0	0.9925
var*env	20	4	-0.0226	2.3581	512	-0	0.9924
var*env	47	4	-0.2001	2.3581	512	-0.1	0.9324
var*env	11	4	-0.2806	2.3581	512	-0.1	0.9053
var*env	36	4	-0.3057	2.3581	512	-0.1	0.8969

var*env	45	4	-0.3184	2.3581	512	-0.1	0.8927
var*env	42	4	-0.4267	2.3581	512	-0.2	0.8565
var*env	32	4	-0.5214	2.3581	512	-0.2	0.8251
var*env	13	4	-0.5429	2.3581	512	-0.2	0.818
var*env	1	4	-0.5494	2.3581	512	-0.2	0.8159
var*env	25	4	-0.5702	2.3581	512	-0.2	0.809
var*env	59	4	-0.5986	2.3581	512	-0.3	0.7997
var*env	22	4	-0.6555	2.3581	512	-0.3	0.7812
var*env	39	4	-0.6821	2.3581	512	-0.3	0.7725
var*env	60	4	-0.7101	2.3581	512	-0.3	0.7634
var*env	27	4	-0.7321	2.3581	512	-0.3	0.7563
var*env	34	4	-0.9275	2.3581	512	-0.4	0.6942
var*env	40	4	-1.0287	2.3581	512	-0.4	0.6628
var*env	3	4	-1.052	2.3581	512	-0.5	0.6557
var*env	61	4	-1.0655	2.3581	512	-0.5	0.6516
var*env	9	4	-1.0714	2.3581	512	-0.5	0.6498
var*env	6	4	-1.1191	2.3581	512	-0.5	0.6353
var*env	18	4	-1.1509	2.3581	512	-0.5	0.6257
var*env	7	4	-1.2333	2.3581	512	-0.5	0.6012
var*env	58	4	-1.2352	2.3581	512	-0.5	0.6006
var*env	62	4	-1.349	2.3581	512	-0.6	0.5675
var*env	55	4	-1.3764	2.3581	512	-0.6	0.5597
var*env	8	4	-1.6182	2.3581	512	-0.7	0.4929
var*env	21	4	-1.7412	2.3581	512	-0.7	0.4606
var*env	38	4	-1.7751	2.3581	512	-0.8	0.4519
var*env	24	4	-1.8083	2.3581	512	-0.8	0.4435
var*env	4	4	-1.9515	2.3581	512	-0.8	0.4083
var*env	52	4	-2.0051	2.3581	512	-0.9	0.3955
var*env	23	4	-2.0131	2.3581	512	-0.9	0.3937
var*env	49	4	-2.0204	2.3581	512	-0.9	0.3919

var*env	49	5	4.0785	2.3581	512	1.73	0.0843
var*env	52	5	3.3532	2.3581	512	1.42	0.1556
var*env	62	5	3.3226	2.3581	512	1.41	0.1594
var*env	34	5	2.2628	2.3581	512	0.96	0.3377
var*env	33	5	2.2124	2.3581	512	0.94	0.3486
var*env	60	5	2.2042	2.3581	512	0.93	0.3503
var*env	32	5	2.0428	2.3581	512	0.87	0.3867
var*env	55	5	1.8476	2.3581	512	0.78	0.4337
var*env	22	5	1.7203	2.3581	512	0.73	0.466
var*env	28	5	1.6361	2.3581	512	0.69	0.4881
var*env	25	5	1.6103	2.3581	512	0.68	0.495
var*env	61	5	1.5458	2.3581	512	0.66	0.5124
var*env	38	5	1.4759	2.3581	512	0.63	0.5317
var*env	24	5	1.308	2.3581	512	0.55	0.5794
var*env	21	5	1.0183	2.3581	512	0.43	0.666
var*env	23	5	1.0157	2.3581	512	0.43	0.6669
var*env	45	5	1.0003	2.3581	512	0.42	0.6716
var*env	50	5	0.9087	2.3581	512	0.39	0.7001
var*env	48	5	0.7043	2.3581	512	0.3	0.7653
var*env	39	5	0.6635	2.3581	512	0.28	0.7785
var*env	43	5	0.6002	2.3581	512	0.25	0.7992
var*env	20	5	0.542	2.3581	512	0.23	0.8183
var*env	58	5	0.5278	2.3581	512	0.22	0.823
var*env	54	5	0.4118	2.3581	512	0.17	0.8614
var*env	41	5	0.3998	2.3581	512	0.17	0.8654
var*env	26	5	0.3164	2.3581	512	0.13	0.8933
var*env	18	5	0.2687	2.3581	512	0.11	0.9093
var*env	37	5	0.2479	2.3581	512	0.11	0.9163
var*env	36	5	0.1983	2.3581	512	0.08	0.933
var*env	29	5	0.1356	2.3581	512	0.06	0.9542
var*env	27	5	0.1019	2.3581	512	0.04	0.9656
var*env	47	5	0.08847	2.3581	512	0.04	0.9701
var*env	19	5	0.00652	2.3581	512	0	0.9978
var*env	40	5	-0.0265	2.3581	512	-0	0.991
var*env	53	5	-0.0343	2.3581	512	-0	0.9884
var*env	7	5	-0.056	2.3581	512	-0	0.9811
var*env	64	5	-0.0615	2.3581	512	-0	0.9792
var*env	30	5	-0.0916	2.3581	512	-0	0.969
var*env	31	5	-0.2588	2.3581	512	-0.1	0.9126
var*env	42	5	-0.5152	2.3581	512	-0.2	0.8271
var*env	56	5	-0.5979	2.3581	512	-0.3	0.7999
var*env	46	5	-0.6964	2.3581	512	-0.3	0.7679
var*env	59	5	-0.7275	2.3581	512	-0.3	0.7578
var*env	4	5	-0.9358	2.3581	512	-0.4	0.6916
var*env	17	5	-1.0015	2.3581	512	-0.4	0.6712
var*env	57	5	-1.0144	2.3581	512	-0.4	0.6672
var*env	11	5	-1.1164	2.3581	512	-0.5	0.6361
var*env	13	5	-1.2373	2.3581	512	-0.5	0.6
var*env	35	5	-1.2428	2.3581	512	-0.5	0.5984
var*env	1	5	-1.3852	2.3581	512	-0.6	0.5572
var*env	51	5	-1.4213	2.3581	512	-0.6	0.547
var*env	14	5	-1.5002	2.3581	512	-0.6	0.5249
var*env	44	5	-1.5785	2.3581	512	-0.7	0.5035
var*env	15	5	-1.6354	2.3581	512	-0.7	0.4883
var*env	16	5	-1.6921	2.3581	512	-0.7	0.4733
var*env	63	5	-1.7324	2.3581	512	-0.7	0.4629
var*env	12	5	-1.8551	2.3581	512	-0.8	0.4318
var*env	6	5	-2.0087	2.3581	512	-0.9	0.3947
var*env	8	5	-2.1241	2.3581	512	-0.9	0.3681
var*env	10	5	-2.275	2.3581	512	-1	0.3351

var*env	3	5	-2.4197	2.3581	512	-1	0.3053
var*env	9	5	-2.6815	2.3581	512	-1.1	0.256
var*env	2	5	-2.8875	2.3581	512	-1.2	0.2213
var*env	5	5	-2.9661	2.3581	512	-1.3	0.209
var*env	54	6	3.0048	2.3581	512	1.27	0.2032
var*env	27	6	2.9776	2.3581	512	1.26	0.2073
var*env	58	6	2.8313	2.3581	512	1.2	0.2304
var*env	24	6	2.3322	2.3581	512	0.99	0.3231
var*env	62	6	2.1048	2.3581	512	0.89	0.3725
var*env	47	6	2.0822	2.3581	512	0.88	0.3776
var*env	28	6	1.94	2.3581	512	0.82	0.4111
var*env	44	6	1.8762	2.3581	512	0.8	0.4266
var*env	23	6	1.5955	2.3581	512	0.68	0.4989
var*env	49	6	1.4064	2.3581	512	0.6	0.5511
var*env	53	6	1.3064	2.3581	512	0.55	0.5798
var*env	33	6	1.2437	2.3581	512	0.53	0.5981
var*env	55	6	1.229	2.3581	512	0.52	0.6024
var*env	60	6	1.2019	2.3581	512	0.51	0.6105
var*env	45	6	1.1493	2.3581	512	0.49	0.6262
var*env	46	6	1.1021	2.3581	512	0.47	0.6404
var*env	20	6	1.0007	2.3581	512	0.42	0.6715
var*env	21	6	0.8037	2.3581	512	0.34	0.7334
var*env	29	6	0.803	2.3581	512	0.34	0.7336
var*env	36	6	0.7243	2.3581	512	0.31	0.7588
var*env	61	6	0.6512	2.3581	512	0.28	0.7825
var*env	41	6	0.6498	2.3581	512	0.28	0.783
var*env	38	6	0.5948	2.3581	512	0.25	0.801
var*env	56	6	0.5879	2.3581	512	0.25	0.8032
var*env	26	6	0.5865	2.3581	512	0.25	0.8037
var*env	51	6	0.5119	2.3581	512	0.22	0.8282

var*env	64	6	0.4107	2.3581	512	0.17	0.8618
var*env	15	6	0.2708	2.3581	512	0.11	0.9086
var*env	35	6	0.2662	2.3581	512	0.11	0.9102
var*env	42	6	0.1118	2.3581	512	0.05	0.9622
var*env	19	6	0.09489	2.3581	512	0.04	0.9679
var*env	52	6	-0.0325	2.3581	512	-0	0.989
var*env	17	6	-0.0715	2.3581	512	-0	0.9758
var*env	50	6	-0.1543	2.3581	512	-0.1	0.9479
var*env	39	6	-0.184	2.3581	512	-0.1	0.9378
var*env	37	6	-0.2158	2.3581	512	-0.1	0.9271
var*env	25	6	-0.2269	2.3581	512	-0.1	0.9234
var*env	22	6	-0.2583	2.3581	512	-0.1	0.9128
var*env	40	6	-0.3219	2.3581	512	-0.1	0.8915
var*env	34	6	-0.3419	2.3581	512	-0.1	0.8848
var*env	63	6	-0.3783	2.3581	512	-0.2	0.8726
var*env	14	6	-0.6106	2.3581	512	-0.3	0.7958
var*env	13	6	-0.6709	2.3581	512	-0.3	0.7761
var*env	48	6	-0.6953	2.3581	512	-0.3	0.7682
var*env	2	6	-0.7456	2.3581	512	-0.3	0.752
var*env	3	6	-0.7559	2.3581	512	-0.3	0.7487
var*env	9	6	-1.0917	2.3581	512	-0.5	0.6436
var*env	4	6	-1.1167	2.3581	512	-0.5	0.636
var*env	30	6	-1.1882	2.3581	512	-0.5	0.6145
var*env	32	6	-1.3362	2.3581	512	-0.6	0.5712
var*env	16	6	-1.422	2.3581	512	-0.6	0.5468
var*env	10	6	-1.4393	2.3581	512	-0.6	0.5419
var*env	31	6	-1.517	2.3581	512	-0.6	0.5203
var*env	6	6	-1.5837	2.3581	512	-0.7	0.5021
var*env	5	6	-1.713	2.3581	512	-0.7	0.4679
var*env	12	6	-1.76	2.3581	512	-0.8	0.4558

var*env	1	6	-1.7816	2.3581	512	-0.8	0.4503
var*env	59	6	-1.9116	2.3581	512	-0.8	0.4179
var*env	8	6	-1.9415	2.3581	512	-0.8	0.4107
var*env	57	6	-2.0167	2.3581	512	-0.9	0.3928
var*env	43	6	-2.0718	2.3581	512	-0.9	0.38
var*env	7	6	-2.3375	2.3581	512	-1	0.322
var*env	18	6	-2.6053	2.3581	512	-1.1	0.2697
var*env	11	6	-2.9536	2.3581	512	-1.3	0.2109
var*env	48	7	1.9727	2.3581	512	0.84	0.4032
var*env	17	7	1.8693	2.3581	512	0.79	0.4283
var*env	41	7	1.6076	2.3581	512	0.68	0.4957
var*env	4	7	1.5714	2.3581	512	0.67	0.5054
var*env	61	7	1.3061	2.3581	512	0.55	0.5799
var*env	5	7	1.2849	2.3581	512	0.54	0.5861
var*env	19	7	1.2211	2.3581	512	0.52	0.6048
var*env	15	7	1.168	2.3581	512	0.5	0.6206
var*env	45	7	1.0635	2.3581	512	0.45	0.6522
var*env	2	7	0.9461	2.3581	512	0.4	0.6884
var*env	23	7	0.9442	2.3581	512	0.4	0.689
var*env	26	7	0.8913	2.3581	512	0.38	0.7056
var*env	25	7	0.8791	2.3581	512	0.37	0.7095
var*env	59	7	0.8775	2.3581	512	0.37	0.7099
var*env	3	7	0.8483	2.3581	512	0.36	0.7192
var*env	18	7	0.8167	2.3581	512	0.35	0.7292
var*env	38	7	0.7581	2.3581	512	0.32	0.748
var*env	47	7	0.6297	2.3581	512	0.27	0.7895
var*env	37	7	0.6275	2.3581	512	0.27	0.7902
var*env	14	7	0.6031	2.3581	512	0.26	0.7982
var*env	12	7	0.5174	2.3581	512	0.22	0.8264
var*env	20	7	0.4706	2.3581	512	0.2	0.8419
var*env	8	7	0.4302	2.3581	512	0.18	0.8553
var*env	63	7	0.3237	2.3581	512	0.14	0.8909
var*env	16	7	0.3034	2.3581	512	0.13	0.8977
var*env	44	7	0.2016	2.3581	512	0.09	0.9319
var*env	35	7	0.07271	2.3581	512	0.03	0.9754
var*env	27	7	0.03717	2.3581	512	0.02	0.9874
var*env	30	7	-0.0082	2.3581	512	0	0.9972
var*env	51	7	-0.0788	2.3581	512	-0	0.9733
var*env	13	7	-0.1305	2.3581	512	-0.1	0.9559
var*env	46	7	-0.1618	2.3581	512	-0.1	0.9453
var*env	24	7	-0.1841	2.3581	512	-0.1	0.9378
var*env	58	7	-0.2034	2.3581	512	-0.1	0.9313
var*env	57	7	-0.2375	2.3581	512	-0.1	0.9198
var*env	1	7	-0.2514	2.3581	512	-0.1	0.9151
var*env	34	7	-0.2728	2.3581	512	-0.1	0.908
var*env	43	7	-0.2993	2.3581	512	-0.1	0.899
var*env	6	7	-0.3095	2.3581	512	-0.1	0.8956
var*env	11	7	-0.3193	2.3581	512	-0.1	0.8923
var*env	39	7	-0.3573	2.3581	512	-0.2	0.8796
var*env	56	7	-0.3798	2.3581	512	-0.2	0.8721
var*env	60	7	-0.493	2.3581	512	-0.2	0.8345
var*env	10	7	-0.5017	2.3581	512	-0.2	0.8316
var*env	52	7	-0.5559	2.3581	512	-0.2	0.8137
var*env	49	7	-0.6116	2.3581	512	-0.3	0.7955
var*env	33	7	-0.6128	2.3581	512	-0.3	0.7951
var*env	22	7	-0.6403	2.3581	512	-0.3	0.7861
var*env	50	7	-0.6507	2.3581	512	-0.3	0.7827
var*env	31	7	-0.7679	2.3581	512	-0.3	0.7448
var*env	29	7	-0.7707	2.3581	512	-0.3	0.7439
var*env	54	7	-0.784	2.3581	512	-0.3	0.7397

var*env	62	7	-0.8222	2.3581	512	-0.4	0.7275
var*env	28	7	-0.8658	2.3581	512	-0.4	0.7136
var*env	53	7	-0.9944	2.3581	512	-0.4	0.6734
var*env	9	7	-1.0226	2.3581	512	-0.4	0.6647
var*env	40	7	-1.0405	2.3581	512	-0.4	0.6592
var*env	7	7	-1.0565	2.3581	512	-0.5	0.6543
var*env	64	7	-1.1159	2.3581	512	-0.5	0.6363
var*env	32	7	-1.2199	2.3581	512	-0.5	0.6051
var*env	36	7	-1.4217	2.3581	512	-0.6	0.5468
var*env	55	7	-1.4892	2.3581	512	-0.6	0.528
var*env	21	7	-1.5106	2.3581	512	-0.6	0.5221
var*env	42	7	-2.1015	2.3581	512	-0.9	0.3732
var*env	64	8	2.539	2.3581	512	1.08	0.2821
var*env	8	8	1.9171	2.3581	512	0.81	0.4166
var*env	50	8	1.7249	2.3581	512	0.73	0.4648
var*env	37	8	1.6566	2.3581	512	0.7	0.4827
var*env	13	8	1.2419	2.3581	512	0.53	0.5986
var*env	32	8	1.1759	2.3581	512	0.5	0.6182
var*env	39	8	0.9681	2.3581	512	0.41	0.6816
var*env	26	8	0.9509	2.3581	512	0.4	0.6869
var*env	59	8	0.9371	2.3581	512	0.4	0.6912
var*env	23	8	0.8355	2.3581	512	0.35	0.7233
var*env	46	8	0.8336	2.3581	512	0.35	0.7239
var*env	34	8	0.7765	2.3581	512	0.33	0.7421
var*env	9	8	0.7606	2.3581	512	0.32	0.7472
var*env	15	8	0.6217	2.3581	512	0.26	0.7922
var*env	56	8	0.5752	2.3581	512	0.24	0.8074
var*env	49	8	0.5656	2.3581	512	0.24	0.8105
var*env	20	8	0.5234	2.3581	512	0.22	0.8244
var*env	38	8	0.4407	2.3581	512	0.19	0.8518

var*env	55	8	0.4219	2.3581	512	0.18	0.8581
var*env	25	8	0.4202	2.3581	512	0.18	0.8586
var*env	33	8	0.3759	2.3581	512	0.16	0.8734
var*env	40	8	0.3656	2.3581	512	0.16	0.8768
var*env	4	8	0.3114	2.3581	512	0.13	0.895
var*env	19	8	0.2909	2.3581	512	0.12	0.9019
var*env	5	8	0.2403	2.3581	512	0.1	0.9189
var*env	47	8	0.2045	2.3581	512	0.09	0.9309
var*env	6	8	0.1945	2.3581	512	0.08	0.9343
var*env	57	8	0.1789	2.3581	512	0.08	0.9396
var*env	7	8	0.1207	2.3581	512	0.05	0.9592
var*env	44	8	0.09954	2.3581	512	0.04	0.9663
var*env	10	8	0.00897	2.3581	512	0	0.997
var*env	48	8	-0.0078	2.3581	512	0	0.9974
var*env	2	8	-0.0783	2.3581	512	-0	0.9735
var*env	27	8	-0.112	2.3581	512	-0.1	0.9621
var*env	3	8	-0.1357	2.3581	512	-0.1	0.9541
var*env	11	8	-0.2059	2.3581	512	-0.1	0.9304
var*env	43	8	-0.2128	2.3581	512	-0.1	0.9281
var*env	31	8	-0.2235	2.3581	512	-0.1	0.9245
var*env	51	8	-0.228	2.3581	512	-0.1	0.923
var*env	54	8	-0.2733	2.3581	512	-0.1	0.9078
var*env	52	8	-0.3293	2.3581	512	-0.1	0.889
var*env	16	8	-0.3372	2.3581	512	-0.1	0.8863
var*env	14	8	-0.354	2.3581	512	-0.2	0.8807
var*env	12	8	-0.4262	2.3581	512	-0.2	0.8566
var*env	42	8	-0.4598	2.3581	512	-0.2	0.8455
var*env	41	8	-0.5008	2.3581	512	-0.2	0.8319
var*env	53	8	-0.504	2.3581	512	-0.2	0.8308
var*env	1	8	-0.5352	2.3581	512	-0.2	0.8205

var*env	63	8	-0.6132	2.3581	512	-0.3	0.7949
var*env	29	8	-0.6303	2.3581	512	-0.3	0.7893
var*env	58	8	-0.7027	2.3581	512	-0.3	0.7658
var*env	61	8	-0.7552	2.3581	512	-0.3	0.7489
var*env	22	8	-0.8501	2.3581	512	-0.4	0.7186
var*env	62	8	-0.8569	2.3581	512	-0.4	0.7165
var*env	35	8	-0.8709	2.3581	512	-0.4	0.712
var*env	60	8	-0.9182	2.3581	512	-0.4	0.6972

var*env	36	8	-0.9312	2.3581	512	-0.4	0.6931
var*env	18	8	-0.9483	2.3581	512	-0.4	0.6877
var*env	17	8	-1.0942	2.3581	512	-0.5	0.6428
var*env	45	8	-1.2468	2.3581	512	-0.5	0.5972
var*env	24	8	-1.3095	2.3581	512	-0.6	0.5789
var*env	28	8	-1.6075	2.3581	512	-0.7	0.4957
var*env	30	8	-1.8944	2.3581	512	-0.8	0.4221
var*env	21	8	-2.1243	2.3581	512	-0.9	0.3681

DECLARATION

I, the undersigned, declare that the thesis is my original work, has not been presented for degrees in any other University and all source materials used for the thesis have been duly acknowledged.

Name: Demeke Lakew Workie

Signature: ----------------------------------

Place: Faculty of Science, Addis Ababa University

Date: 12/07/10

This thesis has been submitted for examination with my approval as a University advisor.

Name: Girma Taye (Ph.D.)

Signature: ----------------------------------

Date: 12/07/10