Estimation of Genetic Parameter from Trails Designed as Augmented Block Design.

By:

Tadele Kelbessa

A Thesis Submitted to school of graduate studies Jimma University College of Natural Sciences Department of Statistics in the Partial Fulfillment of the Requirement for the degree of Master of Science in Biostatistics

October, 2016

Jimma, Ethiopia

STATEMENT OF AUTHOR

I 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. I have submitted this thesis to Fulfillment of the Requirement for the degree of Master of Science in Biostatistics. The thesis can be deposited in the university library to be made available to borrowers for reference. I solemnly, declare that I have not so far submitted this thesis to any other institution for that award of any academic degree, diploma or certificate. Brief quotations from this thesis are allowed without requiring special permission if an accurate acknowledgement of the source is made.

Tadele Kelbessa

Date: _____

Signature: _____

September, 2016

Jimma Ethiopia

DEPARTMENT OF STATISTICS, SCHOOL OF GRADUATE STUDIES
JIMMA UNIVERSITY

As thesis research advisors, we here by certify that we have read the thesis prepared by Tadele Kelbessa under our guidance, which is entitled "**Estimation of Genetic Parameter from Trials Designed as Augmented Block Design",** in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including tables and figures are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Agatamudi Lakshmana Rao (PhD)   _____   _____

Advisor   Signature   Date

Kibrealem Sisay (M.Sc.)   _____   _____

Co-advisor   Signature   Date

As the members of the board of examiners of M.Sc. thesis open defense examination, we certify that we have read and evaluated the thesis and examined the candidate. Hence, we recommend that the thesis be accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

_____   _____   _____

Name of Chairman,   Signature   Date

_____   _____   _____

Name of Advisor   Signature   Date

_____   _____   _____

Name of Co-Advisor   Signature   Date

_____   _____   _____

Name of internal examiner   Signature   Date

_____   _____   _____

Name of external examiner   Signature   Date

Acknowledgment

Primarily and foremost I would like to thank the almighty God and Savior, Jesus Christ, for giving me life breath and the strength to complete this thesis. I owe you my life Lord!

Next, I want to thank my main advisor Dr.Agatamudi Lakshmana Rao (Dr), who provided valuable guidance and good teaching throughout my studies. His inspiring and enthusiastic being and accessible help has been a great support in writing up of this paper. My deep gratitude goes to my co-advisor, Mr. Kibrealem Sisay who gives me all the necessary advices and comments regarding my thesis development.

I also would like to extend my sincere appreciation and thanks to Sinana Agricultural research center at Oromia Agricultural research institute to undertake this study with their cooperation and permission in using the data.

Last but not the least my appreciation goes to my beloved mother Meselu Belay, to whom I dedicate this thesis, for her endless support. Also I am grateful to my wife Gelane kebede and my brother Tariku Kelbessa for financial and moral support to complete this work and all those who encouraged me in one way or the other.

DEDICATION

I dedicate this work to my beloved mother for making me who I am today, for her support and for teaching me the value of education.

ABSTRACT

The aim of this study was to obtain unbiased estimates of genetic parameters in the application of trials designed in augmented block design. In addition to this, it was planned to compare the variance and covariance components of the intra-block analysis of an augmented experiments and the estimators of the components associated to genetic effect. For analysis and discussion; the data used consisted four local checks and one hundred eighty eight Durum wheat genotypes evaluated using augmented block design at Sinana Agricultural research center in 2013 cropping season. A model that have been contained check treatments and block as fixed factors and Test treatments as random factor have been selected for the data set based on Akaike's Information Createria evidence. The ratios of standard errors of GLM to that of mixed model using trial designed as ABD were about 2.7 for both ML and REML. These results tell us that a mixed model is more valuable than a GLM to remove the down ward bias of variance of the response variable and the boosted variance of the error terms of GLM. The inter-class correlation coefficient result showed that estimation based on REML techniques best to estimate variance component in linear mixed model for trial designed as augmented block design. Finally, the findings of this study showed that about one hundred one genotypes of durum wheat have the highest mean yield effects than standard check genotypes. Therefore about 101Durum wheat genotypes materials have been recommended for next selection program in similar ecology to Sinana.

Keywords: Augmented design, Linear mixed model, General linear model, Variance component, Genotypes, Maximum likelihood, and |Restricted maximum likelihood, variance-component.

Table of Contents

List of Tables

List of Figures

ACRONYMS

ABD:	Augmented Block Designs
ABIBD:	Augmented Balanced Incomplete Block Designs
ACBD:	Augmented Complete Block Designs
AD:	Augmented Designs
AIC:	Akiake's Information Criteria
AICC:	Akiake's Information Criteria Correction
ANOVA:	Analysis of Variance
APBIBD:	Augmented Partially Balanced Incomplete Block Designs
ARCBD:	Augmented Randomized Complete Block Designs
BLUE:	Best Linear Unbiased Estimator
EBLUE:	Empirically Best Linear Unbiased Estimator
EBLUP:	Empirically Best Linear Unbiased Predictor
EM-REML:	Expectation Maximization Restricted Maximum Likelihood Method
GLM:	General Linear Model
H:	Hessian Matrix
I:	Identity matrix
IBD:	Incomplete Block Designs
IC:	Information Criteria
ICC:	Inter-class Correlation Coefficient
LSD:	Least Significance Difference
ML:	Maximum Likelihood
MME:	Mixed Model Equation
MSE:	Mean Square Error
NLME:	Non Linear Mixed Effect
ODS:	Output Delivery System
OLS:	Ordinary Least Squares
PROC GLM:	General Linear Model Procedure
PROC MIXED:	Mixed Procedure
PROC:	Procedure

RCB:	Randomized Complete Block Designs
RCBD:	Randomized Complete Block Designs
REML:	Restricted Maximum Likelihood Method
REML:	Restricted/Residual Maximum Likelihood
SABRAO:	Society for the Advancement of Breeding Research in Asia and Oceania
SAS:	Statistical Analysis System
VC:	Variance Component
UN:	unstructured

1. Introduction

1.1. Background of the Study

Designs are usually characterized by the nature of grouping of experimental units and the procedure of random allocation of treatments to experimental units. Block designs are useful in experiments requiring eliminations of heterogeneity in one or two direction. As experiments become larger with more genotypes and replications, costs will go up. Federer (1956) developed a series of augmented designs to minimize such costs. Augmented designs provide well come flexibility to large experiments and these designs are useful in agricultural experiments. The augmented designs were proposed by Federer (1956),the designs are a modification of straight forward designs, by adding test treatments that appear only once in the experiment to the set of replicated control treatments. Among the augmented designs, the augmented block design is perhaps the most used, and inferences have been made by means of an intra and inter block analysis.

In genetic resources environment, which is a field in the forefront of biological research, an essential activity is to test or evaluate the new germplasm/ provenance / superior selections (test treatments), etc. with the existing provenance or released varieties (control treatments). According to SABRAO (2013) the problem in this evaluation studies is that the quantity of the genetic material collected from the exploration trips is very limited or cannot be made available since a part of this is to be deposited in Gene Bank.

The available quantity of seed is often not sufficient for replicated trials. Moreover, the number of new germplasm or provenance to be tested is very high (usually about 1000-2000 and sometimes up to 3000 accessions), and it is very difficult to maintain the within block homogeneity. On the other hand; when a new varieties or strains are developed in a plant improvement program, sufficient material is often not available for planting more than one experimental plot or unit of the new variety at a single location; in some cases, it may be undesirable to lay out more than one experimental unit for the treatment under consideration. In some plant breeding investigations even though one plot of a new variety is laid out at a single location, the new variety may be planted at a number of locations, with the standard or check varieties being replicated r times at each location. These experimental situations may also occur in the fields of entomology, pathology, chemistry,

physiology, agronomy and perhaps others for screening experiments on new material and preliminary testing of experiments on promising material. For in many cases such design is used with treatments being selection units sampled from a population. Besides, recovery of information on treatments among blocks (inter-block approach) can potentially improve the estimates, and this is achieved by regarding block effects as random.

Designs recognized as having potential to improve the effectiveness (prediction variance) of breeding in field experiments include the randomized complete block design (RCBD), the augmented block design (ABD), and the incomplete block design (IBD), with their possible variations. ABD is most commonly used in the initial steps of a breeding program when there is still a substantial amount of material to be analyzed and, mainly, when there is little propagation material (and thus, the replication of treatments is difficult or impossible). One advantage of ABD is the ease of establishing the experiments, which is particularly useful for breeding programs, for example (Souza *et al.*; 2006; Peternelli *et al.*,2009).

Various augmented experiment designs have been presented in the literature (Federer, 1955, 1961; Federer and Raghavarao, 1975; Federer, Nair and Raghavarao, 1975; Federer and Wright, 1988). The purpose of this study is to estimate genetic parameters in the trial designed by augmented block designs using information obtained from the random effects and from the distributional effects of the augmented (or new) treatments in the experiment. Since augmented designs are used to include treatments (varieties) for which there is little information and often limited material, these treatment or varietal effects can be considered to be distributed around some mean and with a common variance σ^2_T. Herein we consider that each new treatment is included once in the experiment but this need not be the cases the procedure is easily extendable to take additional replication into account. There are c checks or standard treatments which are used to obtain the experiment design prior to adding the augmented treatments.

The check treatment yields are used to obtain solutions for blocking and check treatment effects. The former are used to adjust the new treatment effects. From the mean square for the new treatments, an estimate of the variance component σ^2_T is obtained and used for adjusting new treatment means for their distributional effects. Adjustment for distributional properties of the random effects makes use of all information from an experiment.

Genetic variances can be underestimated if the intra-block analysis is used in the augmented block design (Bearzoti, 1994). Therefore, the theory of mixed models (Henderson, 1984) could be used to take into account the randomness of the effects of test treatments and/or blocks (Robinson, 1991). The effects of control could be considered fixed in plant breeding, for they are generally standard released varieties.

In plant selection programmed, breeders usually start with a large number of test lines, which come either crossing or through introduction from foreign sources. The number of lines can range even up to hundreds. To conduct a field experiment for such a large population is extremely difficult for a number of reasons, among them that environmental heterogeneity in the field cannot be easily taken in to account. To complicate the matter further, material available for each test line is often limited, sometimes being sufficient for only one replication. Thus designs for variety trials involving large numbers of test lines, for example, lattice square and quasi-factorial designs, all of which require replications, cannot be used; similarly, designs such as chain blocks (Youden and Connor, 1953; Mandel,1954), which require that a substantial number of test lines have at least two replication cannot be applied.

In plant breeding programs, there are cases where controls include new treatments generated through breeding. The test treatments are supposed to be compared against certain controls. Any of the experimental design can be used depending on the number of treatment and stage of breeding programs. In the early stage of selection process, there could be insufficient seed of the new treatments' to undertake replicated experiments or the number of genotypes could be very large to manage in terms of resources. In such cases some plant breeder use single row plots to evaluate the newly developed test treatments and at a certain regular intervals control varieties are planted. The performance of the new treatments is then compared with the performance of the nearest controls subjectively. The standard statistical analysis is not possible as the new genotypes are not replicated but an objective comparison can be made.

The disadvantages of subjective judgment are that the controls are systematically placed and no provision is made to adjust a given measurement for environmental variation from one part of the experiment to another. A better method, when there are many test treatments at the early stage of the breeding programs is to use an experimental design called augmented design that

was developed by Federer (1956) and well illustrated by Federer and Raghavarao (1975). This design is of particular interest in an extensive plant breeding programs.

To circumvent the difficulties arising from non replicated experiments, Federer (1956) proposed a class of design called 'augmented design'. The basic idea is to include control lines for which enough material is available and repeat them several times in a standard design. Each repetition of the control lines is embedded in a block (or incomplete block, or cell, depending on the design used) and test lines are assigned to plots that are not allocated to controls. Estimations of block effects and plot error are done only with respect to control lines. The estimated block effects are used to adjust the observed values of the test lines and the error is used to test the significance of line differences.

Augmented designs are un-replicated designs where field variation may be controlled using several different approaches. Traditional un-replicated designs control local variation using a single replicated control variety distributed often systematically across the field. The approach is flexible and simple to use in that genotypes need not be randomized, visual evaluation is possible and the test genotypes are adjusted using the mean yield of the neighboring controls.

Augmented designs eliminating heterogeneity in one direction are called augmented block designs. Federer and Raghavarao (1975) who obtained augmented designs using RCB design for one-way heterogeneity setting gave a general theory of augmented designs. The estimable contrasts in such designs may be (i) among new varieties (test treatments), (ii) among check varieties (control treatments), and (iii) among all check and new varieties simultaneously. Indeed it may be possible to estimate the contrasts between check and new varieties. In this study we concentrate on augmented block designs which minimize the variance of certain contrasts of the new varieties.

Augmented block experimental designs fall into two categories, complete blocks and incomplete blocks for the check genotypes or treatments. A randomized complete block design (RCBD), with r replicates or blocks, is used for the c check genotypes to start the construction of an augmented randomized block. Then, the r blocks are expanded to include the c checks plus n/r new genotypes in each block. If n is not a multiple of r, then fewer or more new genotypes would appear in some of the blocks. The c checks and n/r new genotypes are randomly allotted

to the experimental units in each block. Genotype numbers are randomly assigned to the new genotypes, but this is not necessary in the early stages of screening since each new genotype is a random event in itself.

1.2. Statement of the Problem

The best strategy of increasing production of crops is by increasing productivity per unit area using improved production technology. On the other hand, reliable improved technologies can only be achieved if appropriate designing and modeling is done by accounting for source of variation. Mixed Models is rapidly becoming a very useful tool for statisticians. As a general paradigm it can be used to handle almost every situation, especially if you extend the Linear Mixed Model to the Generalized Linear Mixed Model case or the Nonlinear Mixed Model case. It's also an area in which a lot of research is being done, because the questions are far from being answered. Advanced computing power is giving us the capability to answer those questions. One important question which, unfortunately, still has no good answer is how to estimate variance component and how to select the covariance structure (Chuck. K., 1983). This study, therefore, has tried to fill the gaps in genetic parameter estimation of trial designed as augmented block design.

Generally, this study has attempted to answer the following basic research questions:
1. Which model is appropriate for analysis of data taken from augmented block design?
2. Which estimation method of variance component is contributing more variability to the variability of the model?
3. Which genotypes have more predicted mean difference than local checks?
4. Whether or not the random factor significantly minimize the error mean square of the response variable.

1.3. Objectives of the Study

1.3.1. General Objective

The general objective of this study is to select appropriate model and compare different estimation methods of variance components of selected model in trial designed by augmented block design.

1.3.2. Specific Objectives

- To compare different parameter estimation methods of variance components for trial designed as ABD.
- To present appropriate model for trial designed by ABD for durum wheat genotypes.
- To select genotypes those have more predicted mean difference than local checks?

1.3.3. 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, multifactor 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 field experimenters, breeders and other researchers to conduct research with efficient way, especially on the trial designed as ABD

Moreover, the result helps researchers as a guideline for indicating possible sources of variation that might occur in research activities. The result of this study also help in increasing crop production by giving awareness for the farmers on the recommended wheat genotypes. Also used for further investigations of experimental and statistical problems related with the applications and uses of an ABD in genetic parameter estimation. 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 study on trials ABD.

2. LITERATURE REVIEW

Because of their masking effects on genotypic expression, genotype x environment interaction and soli heterogeneity are the two most important factors that limit progress from selection (e.g., Jones and Frey, 1960; Schutz and Cockerham, 1966; Fasoulas, 1973). Oftentimes, the result is that genetically poor genotypes may appear as good phenotypes because of a positive effect of environmental factors and genetically good genotypes may produce poor phenotypes due to a negative effect of environmental effects (Keuls and Sieben, 1955). The current repertoire of methodologies for addressing the problem of making successful judgments about genotypic values of plants has resulted from research on field experimentation that led to the development and integration of several key principles.

2.1. Basic Concepts

Fisher (1925) laid down the three basic principles of experimental designs as replication, randomization, and local control. Replication and local control were not new, but assigning treatments to plots at random (subject to the restrictions Imposed by the local control) was a major contribution. It assured that the deviations from expected values used for estimating experimental error were independent. Further, randomization made it possible to measure the relative efficiencies of different experimental designs (Yates, 1950). Yates (1950) considered that randomization was the initial fundamental step in the development of field designs for comparisons among genotypes. Additional major developments were factorial designs and balanced incomplete block designs. Méndez (1971) emphasized, however, that even though randomization of the genotypes in a block gave unbiased estimates of the error variance, it did not control intra-block variation.

Replication refers to the number of different plots in which an entry is sown. As expressed by Federer (1984), replication is an essential tool to diminish the effects of soil heterogeneity in biasing genotypic expression. Blocking (or local control) is a method for reducing the variance of a difference between treatment effects (e.g.. Fisher, 1925; Yates, 1936). It groups plots in such a manner that the variation among plots within each group (or block) is minimal and that among groups (or blocks) is maximal. Blocking should be done with as few groups as possible, because degrees of freedom allocated to blocks (incomplete blocks, rows, or columns) reduce those

remaining for the error variance. Statistical analysis and interpretation is complex for designs with blocking.

Yates (1933) introduced the principles of orthogonality and confounding to experimental designs. Orthogonality is that property of the design which ensures that the different classes of effects to which the experimental material is subject will be capable of direct and separate estimation. Randomized complete block and Latin square designs insure that effects are orthogonal to each other, and they are simple to analyze and provide valid estimates of treatment effects and their variances (Federer, 1984). A simple lattice design, on the other hand is not orthogonal. Non-orthogonality sometimes is introduced into experimental designs by deliberately confounding of blocks and treatments (Yates, 1933, 193b). Yates (1933) indicated that where non-orthogonality is deliberate, the design should be such that a slight rearrangement of the data will reestablish orthogonality. For example, non-orthogonality due to incomplete blocks in a simple lattice design can be removed by treating the design as a randomized complete block design (e.g., Yates, 1939, 1940). Thus, a lattice design can never be less efficient than the standard randomized complete block design because if blocking is ineffective in reducing error variance, the experiment can be analyzed as a randomized complete block design.

Experimental designs are used so that the treatments may be assigned in an organized manner to allow valid statistical analysis to be carried out on the resulting data. Different designs identify different known or suspected sources of variation so that the treatments effects can be evaluated free of extraneous environmental or other influences. Statistical theory also requires that certain conditions be met during the execution of the experiment to permit valid probability statements to be made (Sharma, 2000).

2.2. Augmented Designs

Since Federer published his Article "Augmented (or Hoonuiaku) designs" in 1956, a lot of additional research results have been published. Federer (1961) extensively illustrated arithmetically and algebraically an augmented randomized complete block design and an augmented balanced lattice design. For both designs he considered analyses with and without recovery of inter-block information and provided some discussion on unequally sized incomplete blocks. He pointed out that sufficient replications of controls need to be included to have

sufficient degrees of freedom nineteen for estimating experimental error variance and effects of blocking used to control field heterogeneity (Federer and Raghavarao, 1975). He also gave a precise introduction to some augmented row-column designs (Federer *et al.,* 1975) and to the construction and analysis of augmented lattice square designs (Federer, 2002).Generally, he conclude that the appropriate design for an augmented design is the randomized incomplete block design.

Pinney (1991) has made use of augmented designs for on-farm trials or prototype evaluation trials. He advocated the use of augmented design that minimizes plot number and enables the researcher's and farmer's questions to be answered. It allows the farmer some flexibility to decide what treatments are tested on his/her farms. The technology developed at research station forms the set of core treatments and the farmer defined treatments are called the augmented treatments. He has described a hypothetical alley cropping example to illustrate the applications of augmented designs to participatory on-farm agro-forestry research by taking two core treatments and five augmented treatments. He also conclude that the number of plots per farm depend upon the region, population density and farming system. The more the number of plots available per farm, the more is the scope for within- farm replications or for more the treatment augmentation.

To circumvent the general difficulties that arise from un-replicated experiments, Federer (1956, 1960, 1961, 1972), Searle (.1965), Federer and Raghavarao (1975), Federer et al. (1975), Lin and Poushinsky (1983), and Lin *et al.* (1983) introduced augmented designs. An augmented experimental design is any standard design augmented with additional treatments in the complete block, the incomplete block, the row, the column, etc. (Federer, 1961).

The statistical analyses for augmented designs in which v check genotypes have been replicated r times (or even a particular number of times for each check) and in which v* test genotypes have been shown only once can be done in two equivalent ways (Federer and Raghavarao, 1975): (a) The trial of v+v* genotypes may be analyzed using standard methods for disproportionate numbers in the subclasses; or (b) A statistical analysis can be performed on the check yields only to estimate the error variance. This variance is used for entry comparisons. Federer (1956) pointed out the importance of randomization of checks to obtain unbiased error variances.

Lin and Poushinsky (1983), however, indicated that if the checks are assigned randomly to the plots in a block, their distribution pattern may be irregular and thus may not provide adequate adjustment for soil variation. Since the primary objective of a field test in the early stage of genotype selection is genotypic ranking and not testing of the genotype differences, they concluded that effective adjustment was more important than obtaining an unbiased error variance.

A second problem associated with the random assignment of checks has to do with block shape. These two factors led Lin and Poushinsky (1983) to propose an alternative class of augmented designs whose structure is based on a split-plot design. With these, whole plots can be laid out in any standard design but the arrangement of subplots is always 3x3 with the central plot assigned to a check. Because correlations between plots do not decrease linearly with distance (Briggs and Shebesky, 1968; Le Clerg, 1966), Lin and Poushinsky (1983) concluded that weighted distance measures (Yates, 1936) or random allocation of check plots (e.g., Yates, 1936; Federer, 1958) could not give satisfactory adjustments. Therefore, they suggested that subplots should be square or nearly square so that the distance between the check plot and the eight test plots was relatively uniform.

Lin *et al.* (1983) did a simulation study on a modified augmented Latin square using three adjustment methods. Adjustment by design structure (row and column correction factors) was best when soil variation occurred in one or two directions, but adjustment by regression was best when the variation was multi-directional. Adjustment using the check plots to obtain a fertility index was least satisfactory.

Among the experiment designs in augmented block design are the augmented randomized complete blocks (ARCBD), augmented balanced incomplete block (ABIBD), and augmented partially balanced incomplete blocks (APBIBD). With respect to the augmented or new treatments all these designs are incomplete in that all the new treatments do not appear together in the blocks. Thus, recovery of inter block information is needed for a more efficient analysis. First consider an ARCBD with c checks and n new treatments for a total of $v = c + n$ treatments in r blocks. Let the c check treatments appear once in each of the blocks (note that the c treatments could appear in the proportions $n_1 : n_2 : ... : n_c$ in each of the r blocks and the design

would still be an orthogonal one (See Federer, 1991, ch. 7)). Since the *n* new treatments each occur once in the experiment the observation can only contribute to the new treatment estimate and nothing to block, overall mean or error estimation (Federer and Raghavarao, 1975).

2.3. The use of Checks to Control the Effects of Variability

Several methods for making comparisons among a large number of strains on the basis of checks have been described (e.g. Le Clerg, 1966; Yates, 1936). Sometimes checks are placed at regular intervals over the test site, but a systematic arrangement may give a biased estimate of error variance (Yates, 1936; Federer, 1961). Random assignment of strains may lead to biased comparisons with the checks. The strain performance can be expressed relative to the nearest check plot. Another procedure of adjusting a genotype's performance consists of a relative value computed by dividing the plot yield by a fertility index which is calculated on the basis of regression between two check plots. Several studies have been conducted in which the fertility indexes were used as covariates (Ïownley-Smith and Hurd, 1973; Mak *et al.*, 1978; Rosielle, 1980).

Based upon Smith's coefficient of soil variability Smith (1938), Baker and McKenzie (1967) concluded that the use of systematically arranged check plots as fertility indexes was of questionable value. They found no advantage for check plots unless Smith's coefficient was less than 0.5 in the experiment. Briggs and Shebesky (1968) circulated a questionnaire to wheat breeders around the world and found that their use of check to experimental plots ranged from 1 in 200 plots to 1 in 3. The average was 1 check per 50 plots. Townley-Smith and Hurd (1973) compared the efficiencies of adjusting plot yields via repeated checks and moving averages, and found that the latter method gave the best control of error variance. Mak *et al*. (1978) found that adjustment by either method was superior to no adjustment for grain protein content, but not for grain yield. Both techniques gave similar control of experimental error. A balanced lattice design was superior in error control to either adjustment method, however. Rosielle (1980) studied the relative efficiencies of lattice designs, check-plot designs, and moving averages for error variance control in wheat trials. Lattice and check-plot covariance analyses were equally efficient and superior to randomized complete block analyses, but they were only slightly more efficient than moving average covariance analyses.

Schutz and Cockerham (1966) and Yates (1936) concluded that inclusion of check plots for adjusting for block effects was not justified. Another method, discussed by Yates (1936), consisted of dividing the experimental genotypes into sets with each set being sown with one or more checks in several replications of a randomized complete block or Latin square designs. Adjustments of the experimental strains were done by subtracting the mean of the checks for that group. This author recognized that adjustment via check means was unlikely to be as efficient as using an ordinary randomized complete block design.

2.4. Genotype X Environment Interaction

Dudley and Moll (1969) defined the genotype x environment (GxE) interaction variance as being due to failure of genotypes to produce similarly under different environments. The effect of GxE interaction in masking genotypic expression has been amply recognized by plant breeders who attempt to reduce its importance by testing genotypes over many environments. Comstock and Moll (1963) showed statistically how large G x E interactions limits progress from selection.

Several methods have been proposed to solve the problems created by GxE interaction. Horner and Frey (1957) stratified Iowa into homogeneous areas (i.e., they used the blocking principle) to control the variety x location interaction. This approach has been used by McCain and Schultz (1959) and Liang *et al.* (1966) also. The use of regression analysis of genotypic performance on environments to describe and compare genotypic performance was proposed by Yates and Cochran (1938) who subdivided the GxE interaction into sums of squares due to regression and deviations from regression. Finlay and Wilkinson (1963) used this regression approach and defined the regression coefficient as a parameter of adaptability. The deviations from regression provide a measure of stability (Eberhart and Russell, 1966).

Plaited and Peterson (1959) proposed measuring the stability of individual cultivars via combined analyses of variance for all possible pair of genotypes in a set. Wricke (1962) proposed a measure called equivalence and Shukla (1972) computed a stability variance for each variety. Bilbro and Ray (1976) proposed the coefficient of determination as an independent unit criterion for measuring stability of a genotype. Variance components from experiments conducted over environments have been used to determine optimum resource allocation. Sprague and Federer (1951) and Pederson and Rathjen (1981) found that un-replicated trials sown at

several locations and in several years provided the best resource use for making genetic gain. Schutz and Cockerham (1906) suggested that experimental design for control of intra-site error may be relatively unimportant if GxE interaction is large, and Rosielle (1980) found that differences in efficiencies of designs based on intra-site data had little meaning for combined analyses. He concluded that genotype x environment interaction was more important than intra-site error in limiting progress from selection.

2.5. 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.

3. Materials and Methods of Analysis

3.1. Data

3.1.1. Study Area

The data that have been used in this study was taken from the experiments conducted at Sinana Agricultural research center under the Oromia Agricultural Research institute. Sinana is found in Bale Zone of Oromia Region at 463km from Finfine (Addis Ababa). Its geographical location is 07^0 07 ^{N} latitude and $40^0 10^{'}$ E longitude. The elevation of the center is 2400 m.a.s.l with topography of gentle conductive for agricultural production system under rain-fed in the present climatic conditions.

3.1.2. Experimental Materials and Design

The data in this study were taken from experiment have been conducted at Sinana Agricultural research center under Oromia Agricultural Research Institute in the eastern part of Ethiopia during the 2013 G.C. cropping season. The experimental materials consists one hundred eighty eight Durum wheat genotypes evaluated at Sinana (on-station) in 2013 cropping season with the objective of selecting high yielding and disease tolerant durum wheat genotypes. It was carried out by Augmented block design with four local checks (Toltu, Dire, Ejersa and Bakalcha) replicated five times. (Appendix A of table 1).The experiment was laid out in augmented block Design having two rows of 0.2m spaced and 1m length. Seed rate of 150kg/ha and fertilizer rates of 41/46 (N/P$_2$O$_5$) were used respectively. Data to be collected: DH, DM, Plh (cm), TKW (g), TW (kg/hl), GY (g/m2), Disease data (Sr, Yr, Lr and Septoria).

3.2. Augmented Block Designs

In randomized complete block design each treatment is replicated in each block. The design is suitable for single-factor trials as well as multi-factor trials. The RCB design has many advantages. It offers flexibility in the number of treatments and number of replications. Control treatment may be introduced more than once. The statistical analysis of RCB design is simple and rapid. The principal disadvantage of this layout is when the trial should include a large number of treatments. This leads to large complete blocks which may be heterogeneous. Also selection in breeding deals with large number of genotypes which is difficult to include in a block.

Augmented designs are non-replicated experimental designs that circumvent this problem (Federer, 1956; Federer et al., 1975; Federer and Raghavarao, 1975). Augmented block design is suitable for trials including a large number of populations and when the amount of seed is limited and is only enough for one replicate. Controls are repeated systematically in the experiment to control the environmental heterogeneity.

The principles followed in this design are:

- All controls treatment appears in each block.
- Test treatments appear only once in the whole experiment (un-replicated test).
- Block size is determined by number of block, control treatment and test treatments.
- Number of block is determined by test treatment and control treatment.
- Control treatment used to estimate block effects and provides error term
- Effective, but much of the field is taken up with controls.

The control treatments used as a baseline to compare test treatments and allow a certain degree of extrapolation to the performance of test treatment.

3.3. Genetic Statistical Analysis

The fixed effects were estimated and the random effects predicted using the Mixed Procedure (PROC MIXED) of the computer system Statistical Analysis System (SAS) version 9, following a linear mixed model, described by Henderson (1984): $Y = Xr + Za + Wb + e$ Where, Y is the phenotypic data vector, r is the vector of replication effects (fixed) added to the general mean, a is the vector of genetic effects (random), where, $a \sim N(0, G)$ and $G = A\sigma^2 a$, b is the vector of block effects (random), e is the vector of residues (random), where, $e \sim N(0, R)$ and $R = I\sigma^2 e$. The capital letters represent the matrix incidences for these effects, formed by values 0 and 1, which associate the unknown r, a and b with data vector y, respectively.

Vector r contemplates all replications of all places, in other words, the effects of location and replication within locations. In the mixed models focus, G refers to the matrix of genetic covariance between the progenies, designated $A\sigma^2 a$. For A the parentage coefficient was disregarded here, therefore the matrix G was designated $I\sigma^2 a$, where A=I. Thus, $\sigma^2 a$ is equivalent to the genetic variance between progenies. If the estimates of variances of the random effects are known, the fixed effects can be estimated and the random effects predicted simultaneously by the mixed model equation (MME) given by:

$$\begin{bmatrix} \hat{r} \\ \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W \\ W'X & Z'Z + A^{-1}\,\sigma_e^2/\sigma_a^2 & Z'W \\ X'Z & Z'W & W'W + I\,\sigma_e^2/\sigma_b^2 \end{bmatrix} = \begin{bmatrix} Xy' \\ Z'y \\ Wy' \end{bmatrix}$$

For the above solutions, the genetic and non-genetic components of variance we reconsidered unknown, which is a practical reality, and were estimated by the restricted maximum likelihood method (REML). Once the REML is an iterative process, the numerical algorithm known as Expectation Maximization (with alternating steps of expectation and maximization) was used, which characterizes the algorithm as EM-REML. Thus, from arbitrary initial values to $\hat{\sigma}_a^2$(genetic variance) and $\hat{\sigma}_b^2$ (block variance), solutions for $\hat{r}$, $\hat{a}$ and $\hat{b}$ are obtained. These solutions are used to obtain new estimates of the components of variance and so on, until the convergence is reached. The response to selection by REML was predicted by the mean breeding or genetic values of the selected progenies.

3.4. Statistical Models and Model Selection

To maintain generality, the conventional statistical terms "block," "treatment," and "response" are used in this section. Each trial of this study was considered a block, with the positive and negative control treatments replicated at least once in each trial to provide an estimate of block effects. The response was the average weight of the chicks in a given isolator unit. The observed response to the ith treatment in the jth block can be represented as. $Y_{ij} = \mu + \beta_j + \tau_i + \sigma_{ij}$

Where μ denotes the overall mean; β_j denotes the effect of the jth block, τ_i denotes the effect of the ith treatment and σ_{ij} denotes the random variation of the ith treatment in the jth block. Thus, different standard errors are required for comparing two treatments depending on whether they are control or test treatments and whether or not they are repeated within or across blocks.

The augmented randomized complete block design is a special case of the partially balanced incomplete block design (Cochran and Cox, 1957). The repetition of one or more treatments in every block ensures that the design is connected. For connected designs, all the treatment means and linear contrasts of the treatment means are estimable. Yates (1936, 1940) developed the methods for the analysis of incomplete block designs. The choice of analysis depends upon the nature of the block effect; i.e., is the block effect fixed or random? The fixed effects, or intra-block, analysis Yates (1936) describes the observed response using the statistical model.

Y = XB + e Where y is a vector of observations; X is a known model matrix; B is the vector of unknown fixed effects; and e is a random vector with mean vector 0 and variance covariance matrix $\sigma^2 I$, where σ^2 is the variance and I is the identity matrix. Because of the unequal treatment replication, the block and treatment effects are not-orthogonal. Thus, the treatment sum of squares and means must be adjusted for block effects.

Alternatively, if one of effect is considered random, Yates (1940) showed that not all information about the treatments is contained in intra-block analysis and proposed an "inter-block" analysis to recover this information. Later, Graybill (1961) and Zelen (1957) showed how to perform analyses that combined the intra-block and inter-block information. The combined analysis describes the observed response using the statistical model.

Y = XB + Zu + e Where Z is a known design matrix and u is a vector of unknown random effects. This model is known as a "mixed model" because it includes a design matrix for both the fixed and the random effects.

Although the output follows a different format from GLM, the MIXED procedure also provides corrected F tests, adjusted means and their standard errors, and mean comparisons based on the least significant difference, adjusted for the unequal sample sizes. The ESTIMATE and CONTRAST statements are available to produce custom hypothesis tests as in the GLM procedure.

Information theoretic criteria have played a prominent role in model selection for the linear mixed model. Information theoretic criteria are defined as an estimate of the measure of fit of a model to the data. The most common criteria used in mixed models are the Akaike Information Criterion (AIC, Akaike, 1974) and the Bayesian Information Criterion (BIC, Schwarz, 1978).

The AIC is the directed divergence between the true model and candidate model with respect to the true model. The BIC, on the other hand, technically is not a divergence criterion since it does not assume a true model exists. However, it is generally used as an approximation to a measure of directed divergence.

3.4.1. 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. The terminology mixed models is used when there are models for the fixed effects and one or more than one random effect.

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.

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. Here we need to describe and define concepts of fixed and random effects that are applied in mixed effects model. A model is termed a fixed effects model if all the factors in the variety structure are fixed effects. A model is termed a random effect model if all the factors in the variety structure are random.

A model which contains both fixed and random effect factors with error terms is called mixed model or mixed effects model. 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 (SAS, 1996).

3.4.2. The Model

3.4.2.1. 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:

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{1}$$

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:

$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{E(y)} = \mathbf{Y} - \mathbf{X\beta}$ and thus has E ($\boldsymbol{\varepsilon}$) = 0. And the dispersion matrix var ($\boldsymbol{\varepsilon}$) = $\sigma^2{}_\varepsilon I_N$. 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 Xβ. 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:

$$\mathbf{u} = [\boldsymbol{u_1}{}' \, \boldsymbol{u_2}{}' \, _{...} \, \boldsymbol{u_r}{}'] \tag{2}$$

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

$$\mathbf{Y} = \mathbf{X\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon} \tag{3}$$

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 E(Y) = Xβ and E (Y $|$ u) = Xβ + Zu   (4)

As a result, $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{E} (\mathbf{y} | \mathbf{u})$ (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) = Xβ + 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)= Xβ + ZU** would be the expected value of the random variable Y, given that

---

the random variable U has the value u.

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

$$\text{Var}(\boldsymbol{\varepsilon}) = R = \sigma^2_\varepsilon I_N \tag{6}$$

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

$$\text{Var}(u_i) = \sigma^2_i I_{q_i} \; \forall I \tag{7}$$

With q_i being the number of elements in u, i.e., the number of levels of the factor corresponding to u_i that are represented in the data.

$$\text{Cov}(u_i, u_j) = 0, \; \forall \; i \neq j; \tag{8}$$

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

$$\text{Cov}(U, \varepsilon') = 0 \tag{9}$$

From (7-9), the variance-covariance structure of u is:

$$G = Var(U) = \begin{bmatrix} \sigma_1^2 I_{q1} & \cdots & \cdots \\ \vdots & \sigma_2^2 I_{q2} \; \ddots & \vdots \\ .. & \cdots & \sigma_r^2 I_{r2} \end{bmatrix} \tag{10}$$

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 $Z = [Z_1, Z_2, Z_3... Zr]$ gives

$$Y = X\boldsymbol{\beta} + Z\,U + \boldsymbol{\varepsilon} = X\boldsymbol{\beta} + \sum_{i=1}^{r} ZiUi + \boldsymbol{\varepsilon} \tag{11}$$

Hence from (7) to (11)

$$V = Var(y) = ZGZ' + \sigma_\varepsilon^2 I = ZGZ' + R = \sum_{i=1}^{r} \sigma_i^2 Z_i Z_i' + \sigma_\varepsilon^2 I \tag{12}$$

The mixed linear model represented by $Y = X\boldsymbol{\beta} + Z\,u + \boldsymbol{\varepsilon}$ 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.

3.4.2.2. Description of the Model Used in this Study

In the present study we consider the following:

- Y is 208x 1 vectors of yield of wheat data obtained from one hundred eighty eight different varieties and five blocks with four checks that replicate in each block using 188+5x4 = 208.

- X is a 208x11 design matrices relating the fixed effect (intercept, block and treatn) that contains 0's and 1's. The design matrix X is described as:

X_{i1} = intercept of the model (with entry one)

$X_{i2} = \begin{cases} 1, \text{if data is from block 1} \\ 0, \text{otherwise} \end{cases}$

$X_{i3} = \begin{cases} 1, \text{if data is from block 2} \\ 0, \text{otherwise} \end{cases}$

$X_{i4} = \begin{cases} 1, \text{if data is from block 3} \\ 0, \text{otherwise} \end{cases}$

$X_{i5} = \begin{cases} 1, \text{if data is from block 4} \\ 0, \text{otherwise} \end{cases}$

$X_{i6} = \begin{cases} 1, \text{if data is from block 5} \\ 0, \text{otherwise} \end{cases}$

$X_{i7} = \begin{cases} 1, \text{if data is from treatn1} \\ 0, \text{otherwise} \end{cases}$

$X_{i8} = \begin{cases} 1, \text{if data is from treatn2} \\ 0, \text{otherwise} \end{cases}$

.

.

$X_{i11} = \begin{cases} 1, \text{if data is from treatn5} \\ 0, \text{otherwise} \end{cases}$ where i = 1, 2, 3, ..., 208.

β is a 11x1 vector of fixed effect parameters.

Z is a 208x192 design matrices relating random effects (test treatments).

The design matrix of Z is in the form:

$Z_{i1} = \begin{cases} 1, \text{if data is from 43rd IDYN p\# 3} \\ 0, \text{otherwise} \end{cases}$

$$Zi2 = \begin{cases} 1, \textit{if data is } \text{from CD11 DZMS P\# 6} \\ 0, \textit{otherwise} \end{cases}$$

.

.

.

$$Zi190 = \begin{cases} 1, \textit{if data is } \text{for CD11-DZMS P\#719} \\ 0, \textit{otherwise} \end{cases}$$

$$Zi192 = \begin{cases} 1, \textit{if data is } \text{for checks} \\ 0, \textit{otherwise} \end{cases} \quad \text{where i} = 1, 2, 3 \dots 208$$

- **U** is a 192 x1 vector of unobservable random effects variables u (T.treat). The distribution of U is considered to be normal with mean vector zero and variance-covariance matrix G.

- $\boldsymbol{\varepsilon}$ is a 208x1unobservable vector of random residuals. The distribution of $\boldsymbol{\varepsilon}$ is 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 $\mathcal{E}$ are normally distributed with:

$$E \begin{bmatrix} U \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ And } var \begin{bmatrix} U \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

This study has a random variable for the data set. Therefore, it is a mixed model with one random variable and two fixed effects (block and treatn).

3.5. Data Analysis

The MIXED procedure fits a variety of mixed linear models to data and enables you to use these fitted models to make statistical inferences about the data. A mixed linear model is a generalization of the standard linear model used in the GLM procedure, the generalization being that the data are permitted to exhibit correlation and non constant variability. The mixed linear model, therefore, provides with the flexibility of modeling not only the means of the data (as in the standard linear model) but their variances and co variances as well.

Since Gaussian data can be modeled entirely in terms of their means and variances/ co-variances, the two sets of parameters in a mixed linear model actually specify the complete

probability distribution of the data. The parameters of the mean model are referred to as fixed-effects parameters, and the parameters of the variance covariance model are referred to as covariance parameters. The fixed-effects parameters are associated with known explanatory variables, as in the standard linear model. These variables can be either qualitative (as in the traditional analysis of variance) or quantitative (as in standard linear regression). However, the covariance parameters are what distinguish the mixed linear model from the standard linear model.

The PROC MIXED provides a variety of covariance structures. The most common of these structures arises from the use of random-effects parameters, which are additional unknown random variables assumed to impact the variability of the data. The variances of the random-effects parameters, commonly known as variance components, become the covariance parameters for this particular structure.

Once a model has been fit to the data, we can use it to draw statistical inferences via both the fixed-effects and covariance parameters. PROC MIXED computes several different statistics suitable for generating hypothesis tests and confidence intervals.
The validity of these statistics depends upon the mean and variance-covariance model we select, so it is important to choose the model carefully.

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 test treatments are random factor in the dataset, there exists variance component of random factor in addition to the residual variance. In this study there are two components contributing to the total variance of the observations: test treatments well as a residual variance component. In predicting test treatment that gives best yieldable genotype, we use best Linear Unbiased Prediction (BLUP).

3.5.1. Methods used for estimating Variance Components of Parameters

Variance components models are a way to assess the amount of variation in a dependent variable that is associated with one or more random-effects variables. The central output is a variance components table which shows the proportion of variance attributable to a random effects variable's main effect and, optionally, the random variable's interactions with other factors. Random effects variables are categorical variables (factors) whose categories (levels) are conceived as a random sample of all categories.

Variance components analysis usually applies to a mixed effects model that is, one in which there are random and fixed effects, differences in either of which might account for variance in the dependent variable. There must be at least one random effects variable.

Generally model fitting consists of three parts; estimating variance parameters, fixed effects, and random effects. There is an intensive review and discussion of theoretical aspects and application of estimation methods used in mixed effect models (Littell *et al*., 1996). Unlike the situation with the Generalized Linear Model (GLM) estimation in LMM is based on different principles known by the names REML, ML, MIVQUE0, Types I, II and III. While the ML and REML are based on the maximum likelihood estimation approach which requires the assumption that the distribution of the response is normal MIVQUE0 does not require normality. This study utilizes three estimation methods, and these are the Henderson's Method III, ML and REML.

3.5.1.1. Henderson's Method III.

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

The Method III borrows the approach of 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.

In writing a general mixed model equation as y = Xβ + Zu +e we clearly distinguish between fixed effects and random effects, representing them by β and u, respectively. Suppose for the moment that we remove this distinction and combine β and u into a single vector b and write the model equation as Y = Wb + e and can be considered to be a fixed effects model forgetting all about the differences between fixed and random effects, that is by assuming var (u) = 0. Hence the best linear unbiased estimator of Wb is given as BLUE (Wb) = Wb^0 .

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. A good general summary of the derivation of the mixed model equations wheat 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.

$$f(y,u) = f(y|u).f(u) = \frac{\exp\left\{-\frac{1}{2}(y - x\beta - ZU)' R^{-1}(y - x\beta - ZU) + U' G^{-1}U\right\}}{(2\pi)^{0.5(Nq)}|R|^{0.5}|G|^{0.5}}$$

(**13**)

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 $\tilde{\beta}$ and $\tilde{u}$ to denote the solutions

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ X'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ \tilde{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

(14)

3.5.1.2. Maximum Likelihood Method.

Maximum likelihood estimation begins with writing a mathematical expression known as the Likelihood Function of the sample data. Loosely speaking, the likelihood of a set of data is the probability of obtaining that particular set of data, given the chosen probability distribution model. This expression contains the unknown model parameters. The values of these parameters that maximize the sample likelihood are known as the Maximum Likelihood Estimates.

The method of maximum likelihood estimation, developed by R.A. Fisher in the 1920s (Fisher,

---

1925) seems to have been first applied to the estimation of variance components by Crump (1947, 1951). In this and almost all subsequent presentation of this topic, normality is assumed for the error terms and all the random effects, normality with zero means, homogeneous variance of all random effects pertaining to each factor, and all covariance's zero. Within this framework, Herbach (1959) gave careful attention, for balanced data, to the need for maximum likelihood estimators (MLEs) to be nonnegative, this being essential because ML theory demands that maximization be over the parameter space. In describing ML for variance components it is therefore essential to distinguish between solutions of the ML equations and estimators. They are not necessarily the same. Nor are they always the same as ANOVA estimators.

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.

3.5.1.3. The Maximum Likelihood Equations

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

So for our **y** we have **y** ~N_N (**X**$\boldsymbol{\beta}$, $\boldsymbol{V}$) is a function of multivariate normal distribution with parameters in $\boldsymbol{\beta}$ and **V** to be

$$f(y) = \frac{\exp\{0.5(y - X\beta)' V^{-1}(Y - X\beta)\}}{(2\pi)^{0.5N}|V|^{0.5}}$$

(15)

Where N is the length of y and $\lvert V \rvert$ 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 L = L($\boldsymbol{\beta}, \boldsymbol{V} \mid$ X, y), which is the log likelihood function, in equation (19) y 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:l

$$\ell = \log L = \frac{1}{2}\text{Nlog}2\pi - \frac{1}{2}\log|V| - \frac{1}{2}(y - X\beta)'^{V^{-1}}(y - X\beta)$$

(16)

Differentiating (20) with respect to $\boldsymbol{\beta}$ will be denoted as $\mathbf{l}_{\boldsymbol{\beta}}$ so we get

$$\ell\beta = \frac{\partial \ell}{\partial \beta} = X' V^{-1}y - X' V^{-1}X\beta$$

(17)

Now differentiating equation (15) with respect to $\boldsymbol{\sigma_i}^2$ and making use of:
$\frac{\delta V}{\delta \sigma_i^2} = Z_i Z_i' and \frac{\delta \log |V|}{\delta \sigma_i^2} = V^{-1} \frac{\delta v}{\delta \sigma_i^2} V^{-1}$ for i = 0, 1, 2, … , r as a notational convenience is to
define $u_o = \varepsilon$, q_o = N, $Z_o = I_N$ and $\sigma^2_\varepsilon = \sigma_o^2$, leads to

$$\ell\sigma_i^2 = \frac{\delta \ell}{\delta \sigma_i^2} = \frac{-1}{2} tr(V^{-1}Z_iZ_i') + \frac{1}{2}(y - X\beta)' V^{-1}Z_iZ_i' V^{-1}(y - X\beta)$$

(18)

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

$$X' \dot{V}^{-1} X\beta^{\cdot} = X' \dot{V}^{-1} y$$

(19)

And

$$tr(V^{-1}Z_iZ_i') = (y - X\beta)' V^{-1}Z_iZ_i' V^{-1}(y - X\beta)$$

(20)

For i = 0, 1, 2, …, r an algebraically simpler expression for (18) is derived by defining

$$P = V^{-1} - V^{-1}X(X' V^{-1}X)^{-} X' V^{-1}$$

(21)

Then from (17) it is clear that for $\dot{P}$ being P with replaced by $\dot{V}$

$$\dot{V}^{-1}(y - X\dot{\beta}) = \dot{P}y \tag{22}$$

So that the ML equations (18) and (19) are

Where $\text{tr}\left(V^{-1}Z_iZ_i'\right)$ is a column vector and $\dot{V}$ is estimate of V where $\sigma^{\cdot2}$ is used in place of σ^2

3.5.1.4. 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:

$$\sigma_\varepsilon^{2(m+1)} = \frac{[y'(y - X\beta^{0(m)} - Z\tilde{u}^{(m)})]}{N} \tag{23a}$$

$$\sigma_i^{2(m+1)} = \frac{\tilde{u}'^{(m)}\tilde{u}^{(m)} + \sigma_i^{2(m)}tr(w_{ii}^{(m)})}{q_i} = \frac{\tilde{u}'^{(m)}\tilde{u}^{(m)}}{q_{i - tr(w_{ii})^{(m)}}} \tag{23b}$$

Where $W = (I + Z'R^{-1}ZG)^{-1} = W_{ij}$, 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:

$$\sigma_\varepsilon^{2(m+1)} = \frac{[y'(y - X\beta^{(m)} - Zu^{(m)})]}{N} \tag{23a'}$$

$$\sigma_i^{2(m+1)} = \frac{\tilde{u}'^{(m)}\tilde{u}^{(m)} + \sigma_i^{2(m)}tr(w_{ii}^{(m)})}{q_i} = \frac{\tilde{u}'^{(m)}\tilde{u}^{(m)}}{q_{i - tr(w_{ii})^{(m)}}} \tag{23b'}$$

1. *Decide on starting values* $\sigma_\varepsilon^{2(0)}$ *and* $\sigma_i^{2(0)}$, *and set* $m = 0$.

2. Calculate $W^{(m)} = (\sigma_\epsilon^{2(m)}I + Z'ZG^{(m)})^{-1}\sigma_\varepsilon^{2(m)}$, var(u)=G and now solve for $\beta^{(m)}$ and $V^{(m)}$ and then calculate.

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

3. Now calculate $\sigma_\varepsilon^{2(m+1)}$ and either of the expressions for $\sigma_i^{2(m+1)}$.

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 $\tilde{u} = u^{(m+1)}$. Use $\tilde{\sigma}^2$ and $\widetilde{W}$ to calculate the information matrix, $\mathbf{I}(\tilde{\sigma}_{ML}^2)$.

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).

3.5.1.5. Residual Maximum Likelihood Estimation.

The Restricted Maximum Likelihood Method (METHOD=REML) is similar to the maximum likelihood method, but it first separates the likelihood into two parts: one that contains the fixed effects and one that does not (Patterson and Thompson 1971). The procedure uses a Newton-Raphson algorithm, iterating until convergence is reached for the log-likelihood objective function of the portion of the likelihood that does not contain the fixed effects.

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.

3.5.1.5.1. The Bias

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 C'y = C'Xβ + C'Zu such that no term in β is contained. i. e.; C'Xβ = 0 $\forall \beta$ (R1)

This implies C'X = 0, with C'_{r*N} of rank r= r(X)

$$P = V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1} = C(C'V^{-1}C)^{-1}C' \qquad (R2)$$

3.5.1.5.2. REML Equations

Searle et al (2006, pp. 251- 252) give the REML equations which we summarize as follows: We have that y⁓N (Xβ, V) for C'X = 0 so C'y ⁓N (0, C'VC). Using the Maximum Likelihood estimation equation we have that

$$tr(V^{-1}Z_iZ'_i) = y'\dot{P}Z_i\dot{P}_y , \text{ i=1, 2, 3....r} \qquad (R3)$$

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 (23) becomes:

$$tr[(C' \dot{V} C)^{-1} C' Z_i Z_i' C] = y' C (C' \dot{V} C)^{-1} C' Z_i Z_i' C (C' \dot{V} C)^{-1} C' y \qquad \text{(R4)}$$

Where, i=1, 2, 3,....r

And $P = V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} V^{-1} = C(C' VC)^{-1} C'$ so that equation (R4) now

becomes $tr(\dot{P} Z_i Z_i') = y' \dot{P} Z_i Z_i' \dot{P} y$ i=1,2,3...r $\qquad \text{(R5)}$

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)

$$tr(Z_i' \dot{P} Z_j Z_j' \dot{P} Z_i) \dot{\sigma}^2 = y' \dot{P} Z_i Z_i' \dot{P} y, \quad \text{i= 1,2,3,…,r} \qquad \text{(R6)}$$

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 $P = V^{-1} - V^{-1} X (X' V^{-1} X)^{-1} X' V^{-1} = C(C' V^{-1} C)^{-1} C'$

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:

$$\sigma_\varepsilon^{2(m+1)} = \frac{[(y'(y - X\beta^{0(m)} - Z\tilde{u}^{(m)})]}{N-r} \qquad \text{(R7)}$$

$$\sigma_i^{2(m+1)} = \frac{\tilde{u}'^{(m)} \tilde{u}^{(m)} + \sigma_i^{2(m)} tr(T_{ii}^{(m)})}{q_i} = \frac{\tilde{u}'^{(m)} \tilde{u}^{(m)}}{q_{i - tr(T_{ii})^{(m)}}} \qquad \text{(R8)}$$

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

3.5.2. 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 $E \begin{vmatrix} U \\ \varepsilon \end{vmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $Var \begin{vmatrix} U \\ \varepsilon \end{vmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$

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 $R = \sigma_\varepsilon^2 I_N$ where I_N denotes the $N{\times}N$ identity matrix. As it is shown above, 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.

3.5.2. 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 $(Y - X\beta)'V^{-1}(Y - X\beta)$ 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 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:

$$ML: l(G, R) = -\frac{1}{2}\log|V| - \frac{1}{2}r'V^{-1}r - \frac{n}{2}\log(2\pi)$$

$$REML: l_R(G, R) = -\frac{1}{2}\log|V| - \frac{1}{2}\log|X'V^{-1}X| -- \frac{1}{2}r'V^{-1}r - \frac{n-p}{2}\log(2\pi)$$

Where $r = y - X(X'V^{-1}X)^{-1}X'V^{-1}y$ 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 $2H^{-1}$ 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 $\hat{G}$ and $\widehat{R}$,respectively.

3.5.2. 2. Estimating Fixed Effects and Predicting Random Effects

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 $E[BLUE(\hat{\beta})] = \beta$ and $E[BLUP(\widehat{U})] = E(U)$. For the mixed model given by Equation (6), the BLUE of β is:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y \tag{R9}$$

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:

$$\widehat{U} = GZ'V^{-1}(y - X\hat{\beta}) \tag{R10}$$

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 (R7) and (R8) 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),

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ X'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix} \tag{R11}$$

While these expressions may look considerably more complicated than Equations (R7) and (R8), R^{-1} and G^{-1} are trivial to obtain if R and G are diagonal, and hence the sub-matrices in Equation (R9) are much easier to compute than V^{-1}. A second advantage of Equation (R9) 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 $\hat{\beta}$ and $\hat{U}$ is of order $(p + q)$ x $(p + q)$, which is usually considerably less than the dimensional of V_{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 ε are each multivariate normal with no correlations between them. Equation (R9) 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 $\hat{\beta}$ and $\hat{U}$ are BLUE and BLUP, respectively, under general conditions provided the variances are known.

However, the contrary is usually true, so that (R7) and (R8) cannot be used in their basic forms. In most cases, V = ZGZ'+ 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 G ($\boldsymbol{\theta}$) and R ($\boldsymbol{\theta}$) to obtain $\hat{G}$ = G ($\boldsymbol{\theta}$) and $\hat{R}$ = R ($\boldsymbol{\theta}$). Then $\hat{V}$ is computed as $\hat{V} = Z\hat{G}Z' + \hat{R}$. 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 $\hat{V}$ of V must be used in the computation, providing the "estimated" generalized least squares (EGLS) estimate (BLUE) of **β** and BLUP are becomes:

$$\tilde{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y \tag{R7'}$$

And $\tilde{U} = \hat{G}Z'\hat{V}^{-1}(y - X\tilde{\beta})$ $\tag{R8'}$

3.6. 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:

$$AIC = -2\ell + 2q \text{ And}$$
$$BIC = -2\ell + 2q(logN^*) \tag{R15}$$

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.

3.7. Model Checking

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.

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 residuals 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).

3.8. Computer package use in this study

The Statistical Analysis System (SAS) package has been used in this study for the analysis of the data. It is commonly used statistical software package in agricultural research. The use of SAS in statistical analysis is rapidly increasing with the availability of command-driven SAS in personal computers (PC–SAS). It is powerful in data management and flexible in formatting output.

3.9. Ethical Consideration

The research Ethics Review Board of Jimma University has provided an ethical clearance for the study. The data was brought from Oromia Agricultural research institute and to do so the department of statistics asked to write an official co-operation letter to the Oromia agricultural research institute from where data was obtained. In this research, the information obtained from the researchers is secured.

4. RESULT AND DISCUSSION

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. 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 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).

Table 4.1 shows that all models were comparable as convergence criteria met with positive Hessian matrix. If convergence is met and the estimated Hessian matrices are positive definite 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 variable Test treatments has smallest values of AIC and BIC for ML (AIC = 2331.7 and BIC = 2367.6). This model also has the smallest values of AIC and BIC for REML (AIC = 2250.2 and BIC = 2256.7). Therefore, a model that contains all fixed effects (treatn and block) and random factor Test treatments is appropriate to the data set of this study.

Table 4.1: Selected Information Criterion for model selections.

Variables	Methods						-2LogLike Diff ML (REML)
	ML Criteria			REML Criteria			
	-2LogLike	AIC	BIC	-2LogLike	AIC	BIC	
Fixed	2322.1	2338.1	2335.0	2284.2	2290.2	2289.1	
Fixed and random	2309.7	2331.7	2367.6	2246.2	2250.2	2256.7	12.4 (38)

Now the linear mixed model of this study is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{U} + \boldsymbol{\varepsilon}$ can be fitted to the data as:

Y_{208x1}: a vector representing yield of durum wheat in gram per plot.

X_{208x11}: the design matrix of fixed effects.

β_{11x1}: vector parameters of fixed effects.

$Z_{208x192}$: represents the matrix for random effect.

U_{192x1}: represents the random effect parameters of test treatments.

4.2. Model Checking

As discussed in previews section, we used the PP-plot of residual, histogram of residual, goodness-of-fit tests 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 1 shows that the data conforms to the hypothetical normality assumptions. The fact that the plot is scattered around the straight line and does considerable patter indicates that the distribution of the error term and the response variable is normal (linearity of the error term is fulfilled). This implies that, the residuals are appearing to be fairly normally distributed.
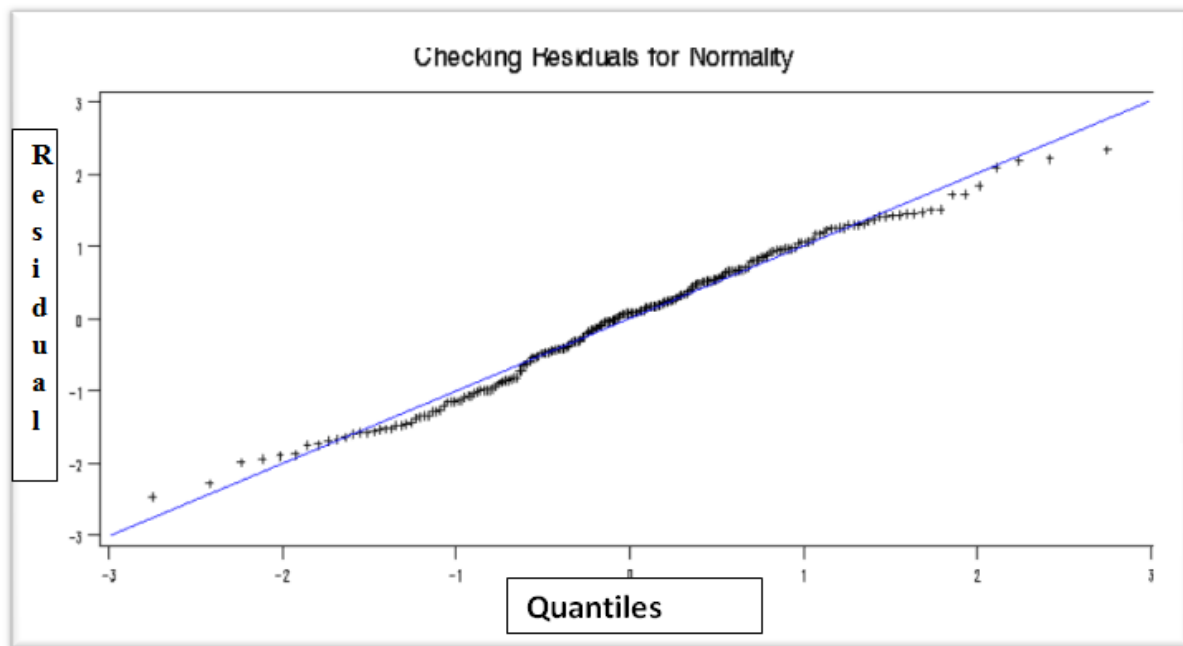
Figure 1: The normal Probability plot of mixed effects for residuals.

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.
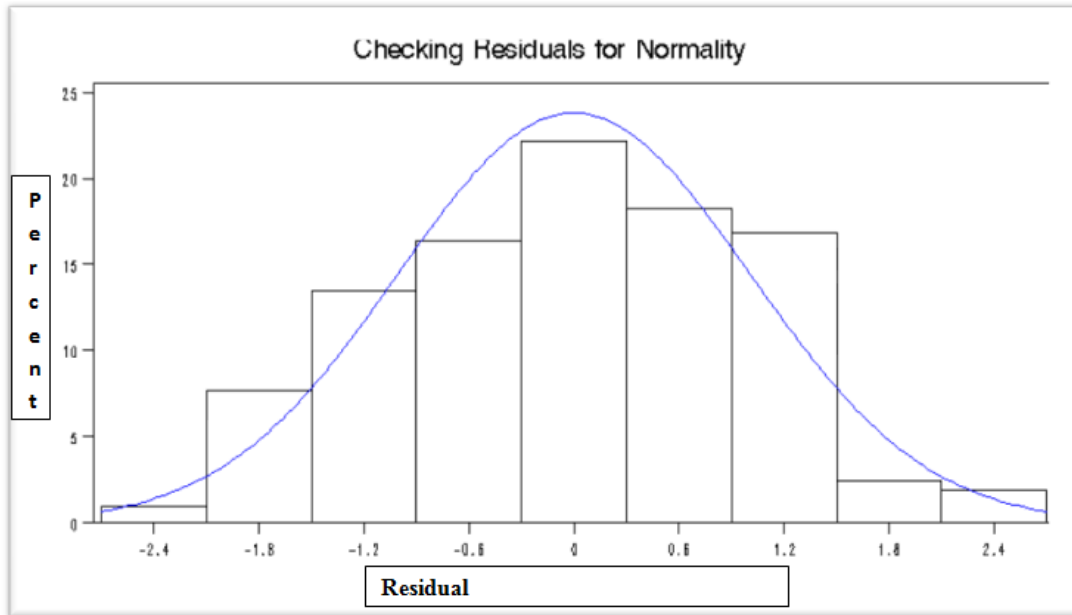
Figure 2: Histogram of mixed effects for Distribution of Residual

The histogram of residual sketched in Figure 2 shows that considerably all observations are at the center and the normal curve is symmetrical about the mean of the residuals. It confirms 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 3 shows that a scatter plots of standardized residual versus the predicted values that includes random effects, $y - X\beta - ZU$ versus $X\beta + ZU$, of the response variable (yield of wheat). 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.
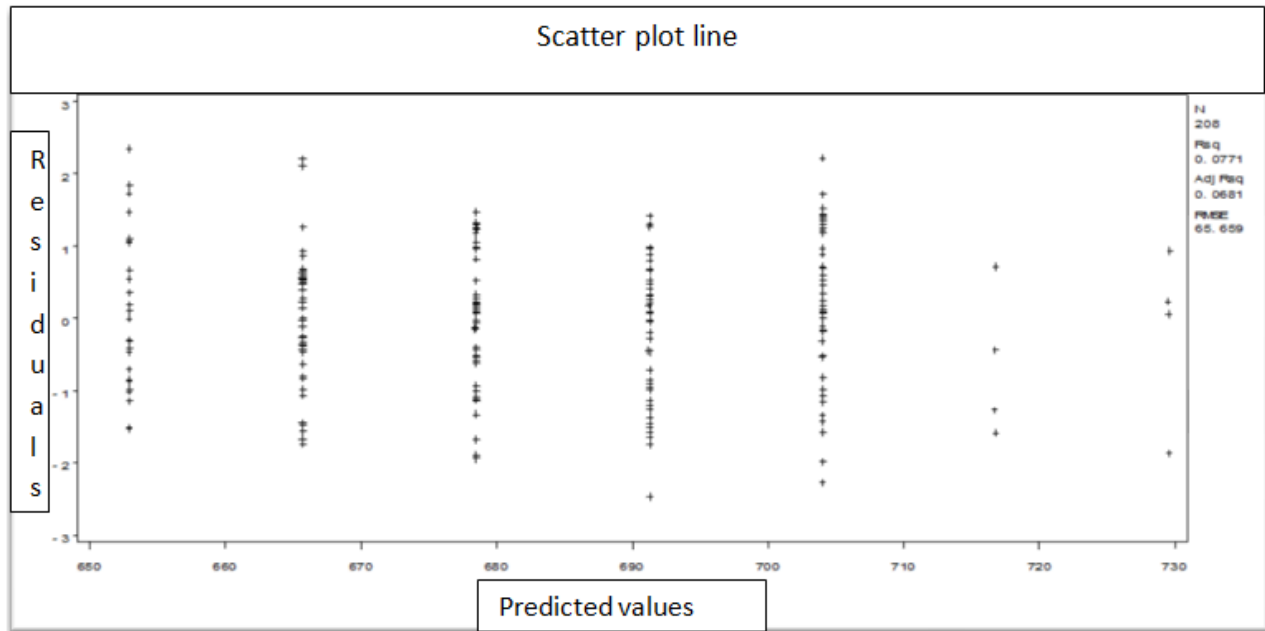
Figure 3: Scatter plot of standardized residuals versus predicted values for mixed effects.

All of the above conclusions are checked by Goodness-of-Fit Tests method for Normal Distribution (normality tests) as the p-value (0.150) of Kolmogorov-Smirnov is greater than level of significant, 0.05 (Table 4.2). 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.

Table 4.2 Tests for Normality of residual by Goodness-of-Fit Tests method

Goodness-of-Fit Tests for Normality distribution				
Test		Statistic	p Value	
Shapiro-Wilk	W	0.987695	Pr < W	0.0696
Kolmogorov-Smirnov	D	0.049846	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.107957	Pr > W-Sq	0.0904
Anderson-Darling	A-Sq	0.737335	Pr > A-Sq	0.0549

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.

4.3. Descriptive statistics

Every good statistical analysis begins with an "ocular test," that is, a good look at the data. Before directly proceeding to graphical representations of the data, it is important to look at the data itself, its structure to determine how best to proceed using the computer. 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.

From Table 4.3, the overall yield of data ranges from 531.9 to 847.6 gram per plot with mean of 682.41gram per plot. The mean yield for check treatments (Toltu, Dire, Ejersa, Bakkalcha) ranged from 725.1, 611.1, 669.7 and 635.5 to 789.1, 772.2, 732.9 and 744.6 with mean of 755.2, 688.94, 692.66 and 679.02 gram per plot respectively. The mean yield of un-replicated treatment (treatn (999)) ranged from 531.9 to 847.6 with mean of 680.11 gram per plot. 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.

Table 4.3: Descriptive statistics for some variables in the study

Variables	N	Mean	Std Dev	Minimum	Maximum
Overall	208	682.40577	68.013956	531.9	847.6
Block 1	44	724.31136	72.612803	556.9	847.6
Block 2	44	673.175	62.903411	531.9	783.7
Block 3	44	674.46364	59.23039	552	773.9
Block 4	44	670.04318	60.813296	552.1	808.2
Block 5	32	665.39688	69.801689	553.9	803.8
treatn189	5	755.2	23.817641	725.1	789.1
treatn190	5	688.94	75.047005	611.1	772.2
treatn191	5	692.66	26.281229	669.7	732.9
treatn192	5	679.02	40.649747	635.5	744.6
Treatn999	188	680.1133	69.185932	531.9	847.6
New 0	20	698.955	58.788161	584	789.1
New 1	188	680.1133	69.185932	531.9	847.6

4.4. Estimation of Variance Components Using Data from trial designed by ABD

The model fitting includes estimation of variance parameters, fixed effects, and random effects. A key assumption in the foregoing analysis for the model that U and ε are normally distributed with zero mean vector and covariance matrix of G and R. The estimates of G and R are calculated using SAS procedure of Henderson method III, ML and REML methods. The fitted model was obtained from model selection procedure as specified in previously, i. e., yield as the dependent variable, block and checks as fixed effects, and test treatments as random effects.

4.4.1. 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 4.4. The variance component estimated for the random part of the model, i.e., for test treatments is 1896.72, whiles the variance component of the error term is 2415.79; it represents the error that remains after the fixed effects and random effects are accounted for.

Table 4.4: Estimation of Variance Components as β and $\mathbf{u}$ are fixed for trial designed as ABD

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	VARCOMP Estimates
block	4	99774	24944	Var (Residual) + 0.9019 Var(T.treat) + Q(block, treatn)	
treatn	4	29203	7300.72	Var(Residual) + 0.0244 Var(T.treat) + Q(treatn)	
T.treat	187	799594	4275.91	Var(Residual) + 0.9807 Var(T.treat)	1896.72
Residual	12	28990	2415.79	Var(Residual)	2415.79

This model is fitted considering that there is no difference between β and $\mathbf{u}$ 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.

The Type I analysis of variance in Table 4.4 consists of a sequential partition of the total sum of squares. The mean square is the sum of squares divided by the degrees of freedom, and the expected mean square is the expected value of the mean square under the mixed model. The "Q" notation in the expected mean squares refers to a quadratic form in parameters of the parenthesized effect. The Type I estimate of the variance components in Table 4.4 result from solving the linear system of equations established by equating the observed mean squares to their expected values.

4. 4.2. Maximum Likelihood Estimation

The "Maximum Likelihood Iterations" in Table 4.5: shows that the Newton-Raphson algorithm required two iterations to converge. The test treatments variance component is 1348.66 and within treatment variance component (Residual) is 1810.21. The estimate in Maximum Likelihood method for residual is smaller than the estimated Variance Component in Type I method. One benefit of using likelihood-based methods is that an approximate covariance matrix is available from the matrix of second derivatives evaluated at the ML solution.

Table 4.5: Estimation of Variance Components using ML for trial designed as ABD.

	Iteration History using ML			Covariance Parameter Estimates using ML				
Iteration	Evaluations	-2 Log Like	Criterion	Cov Parm	Estimate	Standard Error	Z Value	p-value
0	1	2314.5849		T.treat	1346.66	750.45	3.21	0.0007
1	2	2309.72	1.2E-07	Residual	1810.21	597.63	3.03	0.0012
2	1	2309.7199	0					
	Asymptotic Covariance Matrix of Estimates							
Row	Cov Parm	CovP1	CovP2					
1	T.treat	563182	-364466					
2	Residual	-364466	357156					

4. 4.3. Residual Maximum Likelihood Estimation

The "Residual Maximum Likelihood Iterations" in Table 4.6 shows that the REML optimization requires two iterations to converge. The REML estimate variance components of the random effect (test treatments) is 1906.67 and for an error (residual) is 2406.65 (Table 4.6).

Table 4.6: Estimation of Variance Components using REML for trial designed as ABD

Iteration History REML				Covariance Parameter Estimates REML				
Iteratio	Evaluation	-2 Res Log Like	Criterion	Cov Parm	Estimate	Standard Error	Z Value	p-value
0	1	2248.17		T.treat	1906.67	997.35	1.91	0.028
1	2	2246.16	0	Residual	2406.65	884.86	2.72	0.0033

Asymptotic Covariance Matrix of Estimates REML

Row	Cov Parm	CovP1	CovP2
1	T.treat	994702	-787213
2	Residual	-787213	782970

Table 4.7 shows the overall summary estimate for variance components and differences of each other. The difference between estimates of REML and ML for test treatments and the error term are 558.01, and 596.44 ($\sigma^2 u_{REML}$- $\sigma^2 u_{ML}$ and σ^2_εREML- $\sigma^2\varepsilon$ 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.

Table 4.7: The differences of Estimates of Variance Component with Different Techniques for trial designed in ABD

Covariance components	Techniques					
	VARCOMP	ML	REML	REML-VARCOMP	REML-ML	ML-VARCOMP
T.treat	1896.72	1348.66	1906.67	9.95	558.01	-548.06
Residual	2415.79	1810.21	2406.65	-9.14	596.44	596.44

The contribution of 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 tell us the larger the contribution of that random variable for the variation of response variable. The ICC values of test treatments for VARCOMP, ML and REML are 43.9%, 42.7%, and 44.2%, 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.

4.5. Test Statistics of Variance Estimates and Inference for trials designed in ABD

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 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 4.8 Covariance Parameter Estimates for ABD together with Wald Statistic

Covariance Parameter Estimates ML					Covariance Parameter Estimates REML				
Cov Parm	Estimate	Standard Error	Z Value	Pr Z	Cov Parm	Estimate	Standard Error	Z Value	P-value
T.treat	1348.66	750.45	3.21	0.0007	T.treat	1906.67	997.35	2.14	0.028
Residual	1810.21	597.63	3.03	0.0012	Residual	2406.65	884.86	2.72	0.003

Table 4.8 displays covariance parameter estimates together with asymptotic standard errors and the Wald statistic of trial designed in ABD using ML and REML techniques. As we have seen from Table 4.8, the Wald test statistic indicates that the random variable (T.treat) and residuals are significant (p-value <0.05) for ML and REML methods.

As the p-values of the covariance components are small, i. e., <0.05, (Table 4.8) the random variable is 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.

4.6. Advantage of Mixed Effects Model over Fixed Effects Model.

To see the advantage of mixed model over GLM, we need to observe the standard error of mixed model (likelihood estimation) and GLM. The standard errors of residuals from mixed model for trial designed as ABD are 597.63 and 884.86 in ML and REML techniques (Table 4.8) respectively. While the standard errors of residuals from GLM for trial designed as ABD is 2415.79 (Table4.4). The ratios of standard errors of GLM to that of mixed model using trial designed as ABD are about 2.7 for both ML and REML. 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.

4.7. Properties of Fixed & Random Effects for Trial Designed in ABD.

As described in chapter three, the standard method is to solve the mixed model normal equations to obtain estimates of $\boldsymbol{\beta}$ and u. When the inverse of (X'X) does not exist, a generalized inverse can be used in its place. But, in our case X has no full column rank and hence used a generalized inverse of X. Because, the sum of the last 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 over-come 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 $\widehat{G}$ and $\widehat{R}$ are known, $\widehat{\beta}$ is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and $\widehat{u}$ is the best linear unbiased predictor (BLUP) of 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 E $(\widehat{\beta}) = \boldsymbol{\beta}$ and E $(\widehat{u}) = $ E (U). However, G and R are usually unknown and are estimated using one of the above mentioned

(VARCOMP, ML and REML) methods. These estimates, $\hat{G}$ and $\hat{R}$, are therefore simply substituted into the preceding expression to obtain the approximate variance-covariance matrix of $(\hat{\beta} - \boldsymbol{\beta}, \hat{u} - \mathbf{u})$. 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 4.9A, the PROC MIXED procedure in SAS sets zero the estimates of the last level of block (block five). 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 4.9A shows difference mean effect yield of durum wheat genotypes in block five with other block levels. Therefore, the difference mean effect yield of durum wheat genotypes between block five and block one is 58.9145 and so on. Block five is statistical significance difference at alpha level of 0.05 only with blocks one. The difference mean effects of other blocks are statistically not different from zero at alpha level of 0.05.

Table 4.9A: Empirical Best Linear Unbiased Estimates of Fixed Effects (blocks)

Effect	block	Estimate	Standard Error	DF	t Value	P-value
Intercept		665.4	11.4912	203	57.9	<.0001
block	1	58.9145	15.1024	203	3.9	0.0001
block	2	7.7781	15.1024	203	0.52	0.6071
block	3	9.0668	15.1024	203	0.6	0.5489
block	4	4.6463	15.1024	203	0.31	0.7587
block	5	0	.	.	.	.

Table 4.9B: Least Squares Means Estimates for Fixed Effects (blocks).

		Least Squares Means				
Effect	block	Estimate	Standard Error	DF	t Value	p-value
block	1	724.31	9.7998	203	73.91	<.0001
block	2	673.18	9.7998	203	68.69	<.0001
block	3	674.46	9.7998	203	68.82	<.0001
block	4	670.04	9.7998	203	68.37	<.0001
block	5	665.4	11.4912	203	57.9	<.0001

Table 4.9B shows actual estimated means for fixed effects (blocks). It is obtained by summing estimates of the intercept (=block five) and estimates of respective fixed effect levels for blocks. For instance, the actual estimated mean effect for last block is 665.40+0 (=665.40) and the actual estimated mean effect for fourth block is 665.40+4.64 (=670.04). In similar way the other block levels, actual estimated mean effect can be computed.

The predicted mean for last block is 665.40 which is the smallest predicted mean among all blocks. The predicted mean for the second block level is 724.31; that shows the highest predicted mean. The second highest predicted mean has been seen at block3 (=674.46). Table 4.9B also shows that all predicted means for all blocks have p-values less than 0.0001 and were statistically different from zero. This indicates that all blocks have different mean effects on estimating genetic parameters of durum wheat yield.

Table 4.10A shows the difference mean effect yield of durum wheat genetic parameters among checks and un- replicated treatment. The difference mean effect yield of durum wheat genetic parameters between un-replicated treatment (treatn999) and check (189) is 75.0867 and so on. The statistical significance difference is existing at alpha level of 0.05 only between un-replicated treatment (treatn999) and check (189). The other checks and un-replicated treatment (treatn999) are statistically no different mean effect from zero at alpha level of 0.05.

Table 4.10A: Empirical Best Linear Unbiased Estimates of Fixed Effects (treatn)

Effect	Treatn	Estimate	Standard Error	DF	t Value	P-value
			Solution for Fixed Effects			
Intercept		680.11	4.9345	203	137.83	<.0001
treatn	189	75.0867	30.6578	203	2.45	0.0152
treatn	190	8.8267	30.6578	203	0.29	0.7737
treatn	191	12.5467	30.6578	203	0.41	0.6828
treatn	192	-1.0933	30.6578	203	-0.04	0.9716
treatn	999	0	.	.	.	.

Table 4.10B shows actual estimated means for fixed effects (treatn). It is obtained by summing estimates of the intercept (=treatn (999)) and estimates of respective fixed effect levels for treatn. For instance, the actual estimated mean effect for last level of treatn (999) is 680.11+0 (=680.11) and the actual estimated mean effect for treatn (192) is 680.11-1.09 (=679.02). In similar way the other treatn levels, actual estimated mean effect can be computed.

The predicted mean for treatn (bakkalcha (192)) is 679.02 which is the smallest predicted mean among all checks and un-replicated treatments. The predicted mean for the treatn (Toltu (189)) is 755.20; that shows the highest predicted mean. The second highest predicted mean has been seen at treatn (191) is 692.66. This tells us the check treatment Toltu (treatn (189)) has the highest predicted mean and the predicted mean of the un-replicated treatment (treatn (999) = 680.11) is the third highest which greater than the check treatment Bakkalcha (treatn (192) = 679.02). Table 4.10B also shows that all predicted means for all treatn levels have p-values less than 0.0001 and were statistically different from zero. This indicates that all treatn levels have different mean effects.

Table 4.10B: Least Squares Means Estimates for Fixed Effects (treatn).

Effect	treatn	Estimate	Standard Error	DF	t Value	Pr > |t|
			Least Squares Means			
treatn	189	755.2	30.258	203	24.96	<.0001
treatn	190	688.94	30.258	203	22.77	<.0001
treatn	191	692.66	30.258	203	22.89	<.0001
treatn	192	679.02	30.258	203	22.44	<.0001
treatn	999	680.11	4.9345	203	137.83	<.0001

The empirical estimates of random effects with approximate standard errors, the t-test statistic, and p-value are given in Appendix C of Table C2. Empirically best linear unbiased predictor (EBLUP) of yield in Table C2 shows that the overall mean plus average effect of each test treatments to the mean yield of durum wheat.

Appendices C of Table C2 shows that the top one hundred one genotypic materials (test treatments) have been best mean effects to yield production of durum wheat when we compared with standard check variety. The overall empirical estimated predictors of random effects are illustrated in Appendix C of Table C3. From Appendices C of Table C2, varieties CD11-DZMS P#1263 (165), CD11-DZMS P# 1094 (133) and CD11-DZMS P# 1023 (136) have the highest mean effects on the mean yield of durum wheat in decreasing order. But, when we compare each genotypic material (test treatments) with standard check eighty seven genotypic materials (test treatments) produces small empirical estimated predictor to mean yield of durum wheat. For instance, as we see in Appendix C of Table C3, varieties CD11-DZMS P# 1135 (139), CD11-DZMS P# 96 (37) and P#1368 (181) have the smallest mean effects on the mean yield of durum wheat in decreasing order.

4.8. Discussion

This study was used to estimate genetic parameters in the application of trials designed in augmented block design. In addition to this, it was planned to compare the variance and covariance components of the intra-block and inter-block analysis of an augmented experiments. We have seen in this study that the ABD can be analyzed based on the fixed and random effects model. The test treatments where considered as random and other effects as fixed (best to use mixed model procedures for analysis) to see the BLUE and EPLUPS for the realizations of the random and the fixed effects parameters.

The principles and mathematical descriptions of ICCs have been discussed at length in the literature (Eg. Perrett, 2004). Perrett and Higgins (2006) applied a modified form of the ICC to analyze non-replicated greenhouse data. The larger the inter-class correlation coefficient values of the random variable tell us the larger the contribution of that random variable for the variation of response variable. In our study the result of the ICC values high percent were recorded for REML and the difference of ICC with all techniques are too small and estimation based on REML techniques is best to estimate variance components. The results recorded for ICC were comparable with Demeke L. (2010) findings.

In relation to Wald test statistic, a better alternative is the chi-square likelihood ratio test i.e. $\chi 2$. 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 (Self and Liang (1987). The result of Wald test statistic Table 4.8 shows the chi-square statistic for test treatments is 10.304 for ML techniques and 4.5796 for REML methods. The 95 percent of the chi-square distribution with one degree of freedom have critical values of 3.84. Therefore, the test statistic with 5% indicates that the random variable is statistically significant, meaning that the estimate of the

variance of a random effects is above the lower boundary constraint of zero. This implies that a random factor has a contribution for increasing variance of response variable and in turn minimizes error mean squares of response variable. These results also agreed with the privies findings of (Demeke L. ,2010 and Bentler & Bonett's, 1980)

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 tolerate the inclusion random effects. In such kinds of models the variance of the error terms increase very much (Springer, 2000, Temesgen, 2009/10). This illustrates the danger of misusing the statistical models and the rush to irrelevant conclusion. In our data set there are random factors included in the mixed model that contributed their strong share in estimating the variance of the response variable. In this study the ratios of standard errors of GLM to that of mixed model were about 2.7. This tells 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. Our results are comparable to previous studies Demeke L. (2010).

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 (Searle, S. R., Casella, G., and Mcculloch, C.E. (2006). 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 effect the larger contribution to mean yield. In similar way in our study one hundred one treatments have been selected as best mean effects to durum wheat production when compared with standard check variety.

5. Conclusion and Recommendations

5.1. Conclusion

- We have seen in this study that the augmented block design can be analyzed based on the fixed and random effects model. The genetic covariance can be incorporated in to a mixed model; the random effects estimates in the model are typically more efficient than the fixed effects estimates. For the trials designed as ABD the ratios of standard errors of GLM to that of mixed model using trial designed as ABD are about 2.7 for both ML and REML. 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.

- The results of IAC indicate that the models that contend un-replicated treatments as random and other factors as fixed effects have been selected. In our study the REML is the best method of estimation for variance components of genetic parameters for trials designed as augmented block design.

- The results of the present study indicated that about one hundred one genotype materials have been selected as best mean effects to the yield when compared with standard check variety. But, about eight seven of them have been less than standard check variety mean effect on the production of wheat yield.

5.2. 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.

- Wheat researchers, who conduct trials in augmented block design, 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 trials designed as ABD to be REML.
- About 101Durum wheat genotypes materials have been recommended for next selection program in similar ecology to Sinana.

REFERENCES

Akaike, H. (1974). A new look at the statistical model of identification. *IEEE Transaction on Automatic Control*, *19*, 716–723.

Aragaw E. (2011). Construction and analysis of augmented and modified augmented designs.

Arnett, *et al.* 2008. "Patterns of Bat Fatalities at Wind Energy Facilities in North America." Journal of Wildlife Management 72:61-78.

Baker, R. J. and R. I. h. McKenzie.1967. Use of control plots in yield trials. Crop Sci. 7:335-337.

Bearzoti, E (1994). Comparação entre métodos estatísticos na avaliação de clones de batata em um programa de melhoramento. 1994. 128 f. *Dissertação*

Beaton, A. E. (1964), "The Use of Special Matrix Operations in Statistical Calculus," Education Testing Service Research Bulletin, RB-64-51.

Belsley, D. A., Kuh, E. & Welsch, R. E. (1980). Regression Diagnostics: Identifying influential data and sources of collinearity. New York: John Wiley.

Berry, W. D. & Feldman, S. (1985). Multiple regression in practice. London: Sage Publications.

Bilbro, J. D. and L. L. Ray. 1976. Envirnomental stability and adaptation of several cotton cultivars. Crop Sci. 16:821-824.

Bland and Altman, Correlation, Regression and Repeated Data, British Medical Journal, 308, 1994.

Bozdogan, H. (1987), "Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical Extensions," *Psychometrika*, 52, 345–370.

Brlggs, K. G. and L. H. Shebesky. 1968. Implications concerning the frequency of control plots in wheat breeding nurseries. Can. J. Plant Sci. 48:149-153.

Bryk, A.S. and Raudenbush, S.W., *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications, Newbury Park, CA, 2002.

Callanan, T. P., and Harville, D. A. (1991), "Some New Algorithms for Computing Restricted Maximum Likelihood Estimates of Variance Components," Journal of Statistical Computation and Simulation, 38, 239-259.

Cavatassi, K., Richards, M.C., and Heppel, V. (2006). Cereal varieties for the organic and low input grower. p. 147–155.

Chow, S.C. and Shao, J. (1988). A New procedure for the estimation of variance components. *Statist. Prof. Letter 6(5)*: 349-355.

Chuck. K. (1983), Guidelines for Selecting the Covariance Structure in Mixed Model Analysis, COMSYS Information Technology, Inc.

Cochran, W. G., and G. M. Cox, 1957. Experimental Designs. 2nd ed. John Wiley and Sons, New York, NY.

Cochran, W. G. (1939). Long-term agricultural experiments. J. Royal Statistical Soc., B 6:I04-148.

Connor, W. S. (1958). The uniqueness of the triangular association scheme. *Ann. Math. Statist.* **29**, 262-266.

Crossa, J. and P. L. Cornelius (1997). Site regression and shifted multiplicative mode clustering of cultivar trial sites under heterogeneity of error variance. Crop Science 37:406-4I5.

Crump, S.L. (1951). The estimation of variance components in analysis of variances. *Biometrics Bull.,* 2, 7-11.

Demeke L. (2010). Application of linear mixed model to incomplete block designs.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood Estimation from Incomplete- Data via the EM Algorithm" (with discussion), Journal of the Royal Statistical Society, Series B, 39, 1-38.

Dempster, A. P., Selwyn, M. R., Patel C. M., and Roth, A. J. (1984), "Statistical and Computational Aspects of Mixed Model Analysis," Applied Statistics, 33, 203-214.

Douglas, L., Bera, A. K., Jarque, C. M., and Lee, L. F. (1991), Testing the Normality Assumption in Limited Dependent Variable Models, *International Economic*

Eberhart, S. A. and W. A. Russell. 1966. Stability parameters for comparing varieties. Crop Sci. 6:36-40.

Fasoirlas, A. 1973. A new approach to breeding superior yielding varieties. Dept. Genet. Plant

Federer, W. T., 1956. Augmented (or hoonuiaku) designs. Hawaii Plant. Rec. 55:191-208.

Federer, W.T. and R.D. Wolfinger. 2003. Augmented row-column design and trend analysis. p. 291–295. *In* M.S. Kang (ed.) Handbook of formulas and software for plant geneticists and breeders. Food Products Press, Binghamton, NY.

Federer, W.T., R.C. Nair, and D. Raghavarao. 1975. Some augmented Row-column designs. Biometrics 31: 361–373.

Federer, W.T. 1998. Recovery of interblock, intergradient, and intervariety information in incomplete block and lattice rectangle designed experiments. Biometrics 54: 471–481.

Federer, W.T. 2002. Construction and analysis of an augmented lattice square

Fiedler, M. (1971), "Bounds for the Determinant of the Sum of Hermitian Matrices," in Proceedings of the American Mathematical Society, 30, pp. 27-31.

Fessler, J. A., and Hero, A. O. (1994), "Space-Alternating Generalized Expectation-Maximization Algorithm," IEEE Transactions on Signal Processing, 42, 2664-2677.

Finlay, K. W. and G. N. Wilkinson. 1963. The analysis of adaptation in a plant-breeding program. Aust. J. Agric. Res. 14:742-754.plant Breeding, Aristotelian Univ. of thessalonika, Greece, Publ. 3. 114 p.

Foulley, J.-L., and Quaas, R. L. (1995), "Heterogeneous Variance in Gaussian Linear Mixed Models," Genetics Selection Evolution, 27, 211-228.

Fox, J. (1991). Regression diagnostics. London: Sage Publications.

Gaylor, D.W., Lucas, H.L., and Anderson, R.L. (1970), "Calculation of Expected Mean Squares by the Abbreviated Doolittle and Square Root Methods," *Biometrics*, 26, 641 -55.

Girma, T. (2005). Estimation of Optimum Plot Dimension and Replication number for Wheat Experiment in Ethiopia. *African Crop Science Journal,* 1:11-24.

Gleeson, A. C. and Cullis, B. R. 1987. Residual maximum likelihood (REML) estimation of a neighbor model for field experiments. *Biometrics* **43**: 277-288.Federer, W.T. 2005. Augmented split block experiment design. Agron. J. 97: 578–586.

Goodnight, J.H. (1978), *Computing MIVQUE0 Estimates of Variance Components,* S Technical Report R-105. Cary, NC: SAS Institute Inc.

Goldstein, Multilevel Statistical Models, Arnold, 1995

Goodnight, J.H. and Hemmerle, W.J. (1979), "A Simplified Algorithm for the W-Transformation in Variance Component Estimation," *Technometrics*, 21, 265 -268.

Graybill, F. A. (1976), Theory and Applications of the Linear Model, North Scituate, MA: Duxbury

Grondona, M. O. and Cressie, N. A. C. 1991. Using Spatial Considerations in the Analysis of Experiments. *Technometrics*, 33, (4): 381-392. design. Biometrical Journal 44: 241–257.

Gurka, M. J. (2006). Selecting the best linear mixed model under REML. *The American Statistician* 60:19–26.

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.

Hartley, H.O., Rao, J.N.K., and LaMotte, L. (1978), "A Simple Synthesis-Based Method of Variance Component Estimation," *Biometrics*, 34, 233 -244.

Hemmerle, W.J. and Hartley, H.O. (1973), "Computing Maximum Likelihood Estimates for the Mixed AOV Model Using the W-Transformation," *Technometrics*, 15, 819 -831.

Henderson, C.R. (1984). Application of Linear Models in Animal Breeding, University of Guelph

Herbach, L. H. (1959). Properties of Model II type analysis of variance tests, A: optimum nature of the F-test for Model II in the balanced case. *Ann. Math. Statist.* 30, 939-959.

Hicks, C.R. (1973), *Fundamental Concepts in the Design of Experiments*, New York: Holt, Rinehart and Winston, Inc.

Hocking, R.R. (1983), "A Diagnostic Tool for Mixed Models with Applications to Negative Estimates of Variance Components," *Proceedings of the Eighth Annual SAS*

Hocking, R.R. (1984), *The Analysis of Linear Models*, Monterey, CA: Brooks-Cole Publishing

Horn, J., E.B. Arnett, T.H. Kunz. 2008. "Behavioral Responses of Bats to Operating Wind Turbines." Journal of Wildlife Management 72:123-132.

Horner, T. W. and K. J. Frey. 1957. Methods for detecting natural areas for oat varietal recommendations. Agron. J. 49:313-315.

IAR/CIMMY, Addis Ababa, Ethiopia. Henderson, C.R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics,* 31:423-447.

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

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

Jones, K. R. and K. J. Frey. 1960. Heritability percentages and degrees of dominance for quantitative characters in oats. Iowa State J. Scl. 35:49-58.

Jones, R. H. (1990). Serial correlation or random subject affects, *Communications in Stat., Part B–Simulation and Comp.* 19: 1105–1123.

John, P. (1971), Statistical Design and Analysis of Experiments, New York: Macmillan.

Kennedy, W. J., Jr. and Gentle, J. E. (1980), Statistical Computing, New York: Marcel Dekker.

Keuls, M. and J, W. Sleben. 1955. Two statistical problems In plant selection. Euphytica 4:34-44.

Kirk, R. E. (1968), Experimental Design: Procedures for the Behavioral Sciences, Monterey, CA: Brooks-Cole.

Laird, N. (1978). Empirical Bayes methods for two way contingency tables, *Journal of American Statistical Association* 65, 581–590.

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

Le Clerg, E. L. 1966. Significance of experimental design in plant breeding. Pages 243-313, in K. J. Frey, ed. Plant breeding. Iowa State University Press, Ames, lowa.

Leite MSO, Peternelli LA, Barbosa MHP (2006) Effects of plot size on the estimation of genetic parameters in sugarcane families. Crop Breed Appl Biotech 6(1):40–46

Leite MSO, Peternelli LA, Barbosa MHP, Cecon PR, Cruz CD (2009) Sample size for full-sib family evaluation in sugarcane. Pesquisa Agropecuária Bras 44:562–1574

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear Models, *Biometrika* **73**: 13–22.

Lin, C. S. and G. Poushinsky. 1983. A modified augmented design for an early stage of plant selection involving a large number of test lines without replication. Biometrics 39:553-561.

Lindsey, J. K. (1993). *Models for Repeated Measurements*, Oxford University Press, New York.

Littell, Milliken, Stroup and Wolfinger, SAS System for Mixed Models, SAS Institute, 1996

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

Longford, N. T. (1993). *Random Coefficient Models*, Oxford University Press, New York.

Looney, S. W. (1995). How to use tests for univariate normality to assess multivariate normality. American Statistician, 49, 64-70.

Mak, C., B. L. Harvey, and J. 0. Berdahl. 1978. An evaluation of control plots and moving means for error control in barley nurseries. Crop Sci. 18:870-873.

Mandel (1954). Chain block designs with two way eliminations of heterogeneity. *Biometrics, 10, 251-272*

Mariani, B. M. and P. N. Manmana. 1980. On relative efficiency of -incomplete block designs

applied to maize experiments. Maydica 25:1-7.

Masuka BJ, Araus L, Das B, Sonder K, Cairns JE (2012) Phenotyping for abiotic stress tolerance in maize. J Integr Plant Biol 54:238–249

Mendenhall, W. (1968), Introduction to Linear Models and the Design and Analysis of Experiments, Belmont, CA: Duxbury Press.

Méndez, 1. 1971. Estudio de seis metodos alternatives para el usode bloques en la experimentaciôn de campo. Agrociencia 6:3-16.

McCain, F. S. and E. F. Schultz, Jr. 1959. A method for determiningareas for corn varietal recommendations. Agron. J. 51:476-478.

McCulloch and Searle, Generalized, Linear, and Mixed Models, Wiley, 2001

Milliken and Johnson, Analysis of Messy Data Volume I: Designed Experiments, New York, Chapman & Hall, 1989.

Milliken, George A., How to be Successful in Implementing PROC MIXED.

Moehring, J., Williams, E.R. & Piepho, H. Theor Appl Genet (2014) 127: 1049.

NCSS Statistical Software. (2007). NCSS. [Computer Software] Kaysville, UT: Author.

Ofversten, J. (1993). Exact test for Variance Components in Unbalanced Mixed Linear Models. *Biometrics* 49(1): 45-57.

Papadakis J (1984) Advances in the analysis of field experiments. Communicationes dÁcademie dÁthenes 59:326–342

Patterson, H.D. and Thompson, R. (1971), "Recovery of Inter-Block Information When Block Sizes Are Unequal," *Biometrika*, 58, 545 -554.

Pearson, E.S. (1932). Discussion of paper by B.H. Wilsdon, *Supplement of the Journal of the Royal Statistical Society* 1, 200–202.

Pederson, D. G. and A. J. Rathjen. 1981. Choosing trial sites to maximize selection response for grain yield in spring wheat. Aust. J. Agric. Res. 32:411-424.

Perrett, J.J. 2004. Using prior information on the intraclass correlation coefficient to analyze data from unreplicated and under-replicated experiments. Doctoral Dissertation, Kansas State University, Manhattan, KS.

Perrett, J.J., and J. Higgins. 2006. A method for analyzing unreplicated agricultural experiments. Crop Sci. 46:2482–2485.

Plaisted, R. L. and L. C. Peterson. 1959. A technique for evaluating the ability of selections to yield consistently in different locations or seasons. Am. Potato J. 36:381-385.

Press.Hocking, R. R. (1984), Analysis of Linear Models, Monterey, CA: Brooks-Cole.

Rao, C.R. (1971), "Minimum Variance Quadratic Unbiased Estimation of Variance Components," *Journal of Multivariate Analysis*, 1, 445 -456.

Rao, C.R. (1972), "Estimation of Variance and Covariance Components in Linear Models," *Journal of the American Statistical Association*, 67, 112 -115.

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.

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

Romney, D., Kaitho, R., Biwott, J., Wambugu, M., Chege, L., Omore, A., Staal S., Wanjohi, P. & Thorpe, W. (2000).

Rosielle, A. A. Comparison of lattice designs, check plots, and moving means in wheat breeding trials. Euphytica, v. 29, p. 129-133, 1980.

SABRAO Journal of Breeding and Genetics 45 (1) 1-5, 2013

SAS Institute Inc. 2003. SAS®. Version 9.1 [computer program]. SAS Institute Inc., Cary, N.C.

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

SAS Institute. 2004. *SAS/STAT 9.1 User's Guide*. Cary, NC: SAS Institute.

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

Searle, S.R., Casella, G., and McCulloch, C.E. (1992), *Variance Components,* New York: John Wiley and Sons, Inc.

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

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

Schroeder, L. D., Sjoquist, D. L. & Stephan, P. E. (1986). Understanding regression analysis. Beverly Hills, CA: Sage Publications.

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

Schabenberger, Mixed Model Tools in SAS/Stat®, ASA Statistical Consulting Section

      undtable Conference Call, September, 2004.

Schutz, W. M. and C. C. Cockerham. 1966. The effect of field blocking on gain from selection.

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

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality.

      Biometrika, 52, 591-611.

Sharma KC, Sharma RK, Singhania DL, Singh D (2003) Variation and character association in

      fodder yield and related traits in pearl millet (Pennisetum glaucum (L.) R. Br.). Indian

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.

Shukla, G. K. 1972. Some statistical aspects of partitioning genotype environmental

      components of variablility. Heredity 29:237-245.

Smith, H.F. An empirical law describing heterogeneity in the yields of agricultural crops.

      Journal of Agricultural Science, v.28, p.1-23, 1938.

Temesgn, Z. (2009/10). Special Topic: Mixed Model. Lecture notes for AAU second year

      Statistics M.Sc students. *Review*, 25, 563–578.

Townley-Smith, T. F. and W. A. Hurd. 1973. Use of moving means in wheat yield trials. Can.

      J. Plant Sci. 53:447-450.

      Biometrics 22:843-863.

Verbeke and Molenberghs, Linear Models for Longitudinal Data, Springer, 2000

Users Group International Conference, Cary, NC: SAS Institute Inc., 8, 711 -716.

Wilkinson, L, & Task Force on Statistical Inference. (1999). Statistical methods in psychology

      journals: Guidelines and explanations. American Psychologist, 54, 594-604.

Wolfinger, R. D., 1996: Heterogeneous variance-covariance structures for repeated measures.

      J. Agric. Biol. Environ. Stat. 1, 205–230.

Wolfinger, R.D., W.T. Federer, and O. Cordero-Brana. 1997. SAS PROC GLM and PROC

      MIXED for recovering blocking and variety information in augmented designs.

      Agronomy Journal 89: 856-859

Wonnacott, T. H. & Wonnacott, R. J. (1981). Regression: A second course in statistics.

      New York: Wiley

Wricke, G. 1962. Uber eine Methode Ztr Erfassung der okologischen Streubreite in
   Feldversuchen. A. Pflanzenzucht. 47:92-96.

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.

Yates, F., and W.G. Cochran. 1938. The analysis of groups of experiments. J. Agric. Sci.
   28:556–580.

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

Appendix A

Table A1: Durum wheat Genotypes in the study

Genotype	Codes	Genotype	Codes	Genotype	Codes
43rd IDYN p#3	1	CD11-DZMS P# 83	34	m6 srr SN P# 204	67
CD11DZMS P#6	2	43 IDSN P#85	35	m6 srr SN P# 211	68
43rd IDYN p#11	3	43 IDSN P#93	36	CD11-DZMS P# 222	69
43th IDYN*H p#15	4	CD11-DZMS P# 96	37	CD11-DZMS P# 251	70
CD11DZMS P#14	5	43 IDSN P#96	38	Ejersa	191
m6 srr SN P#16	6	43 IDSN P#97	39	CD11-DZMS P# 283	71
43rd IDYN p#3	7	43 IDSN P#100	40	CD11-DZMS P# 267	72
43th IDYN P#15	8	Bakkalcha1	192	CD11-DZMS P# 281	73
m6 srr SN P#18	9	43 IDSN P#101	41	CD11-DZMS P# 289	74
43rd IDYN P#20	10	43 IDSN P#109	42	CD11-DZMS P# 309	75
Toltu	189	CD11-DZMS P# 108	43	CD11-DZMS P# 311	76
43 IDSN P#22	11	CD11-DZMS P# 110	44	CD11-DZMS P# 317	77
CD11-DZMS P#21	12	43 IDSN P#113	45	CD11-DZMS P# 318	78
43 IDYN*H p# 22	13	43 IDSN P#114	46	CD11-DZMS P# 321	79
43 IDSN P#23	14	43 IDSN P#120	47	CD11-DZMS P# 352	80
43 IDYN*H p# 25	15	43 IDSN P#122	48	Bakkalcha2	192
CD11-DZMS P#27	16	CD11-DZMS P# 123	49	CD11-DZMS P# 353	81
CD11-DZMS P#30	17	43 IDSN P#129	50	CD11-DZMS P# 354	82
43rd IDYN P#39	18	Toltu	189	CD11-DZMS P# 355	83
43 IDYN*H p# 35	19	43 IDSN P#133	51	CD11-DZMS P# 356	84
43 IDSN P#36	20	43 IDSN P#135	52	CD11-DZMS P# 358	85
Dire	190	m6 srr SN P# 136	53	CD11-DZMS P# 359	86
43rd IDYN P#36	21	CD11- DZms p#137	54	CD11-DZMS P# 360	87
43 IDSN P#32	22	CD11- DZms p#240	55	CD11-DZMS P# 361	88
43 IDYN *H P#37	23	43 IDSN P# 143	56	CD11-DZMS P# 364	89
CD11-DZMS P#37	24	CD11-DZMS P# 146	57	CD11-DZMS P# 366	90
43 IDYN *H P#34	25	CD11-DZMS P# 150	58	Toltu	189
43 IDYN *H P#40	26	m6 srr SN P# 152	59	CD11-DZMS P# 367	91
CD11-DZMS P#42	27	m6 srr SN P# 153	60	CD11-DZMS P# 370	92
CD11-DZMS P#48	28	Dire	190	CD11-DZMS P# 393	93
P# 57	29	43 IDSN P# 153	61	CD11-DZMS P# 394	94
CD11-DZMS P# 64	30	m6 srr SN P# 157	62	CD11-DZMS P# 395	95
Ejersa	191	CD11-DZMS P# 163	63	CD11-DZMS P# 403	96
p#60	31	CD11-DZMS P# 167	64	CD11-DZMS P# 404	97
43 IDSN P#57	32	CD11-DZMS P# 168	65	CD11-DZMS P# 406	98
43 IDSN P#71	33	CD11-DZMS P# 169	66	CD11-DZMS P# 408	99

CD11-DZMS P# 415	100	CD11-DZMS P# 744	120	CD11-DZMS P# 1137	140
Dire	190	Bakkalcha3	192	Dire	190
CD11-DZMS P# 419	101	CD11-DZMS P# 743	121	CD11-DZMS P#1136	141
CD11-DZMS P# 454	102	CD11-DZMS P# 769	122	CD11-DZMS P#1138	142
CD11-DZMS P# 489	103	CD11-DZMS P# 372	123	CD11-DZMS P#1139	143
CD11-DZMS P# 511	104	CD11-DZMS P# 777	124	CD11-DZMS P#1158	144
CD11-DZMS P# 512	105	CD11-DZMS P# 782	125	CD11-DZMS P#1163	145
CD11-DZMS P# 513	105	CD11-DZMS P# 785	126	CD11-DZMS P#1145	146
CD11-DZMS P# 549	107	CD11-DZMS P# 847	127	CD11-DZMS P#1162	147
CD11-DZMS P# 548	108	CD11-DZMS P# 848	128	CD11-DZMS P#1167	148
CD11-DZMS P# 550	109	CD11-DZMS P# 856	129	CD11-DZMS P#1164	149
CD11-DZMS P# 646	110	CD11-DZMS P# 866	130	CD11-DZMS P#1147	150
Ejersa	191	Toltu	189	Ejersa	191
CD11-DZMS P# 653	111	CD11-DZMS P# 875	131	CD11-DZMS P#1170	151
CD11-DZMS P# 663	112	CD11-DZMS P# 949	132	CD11-DZMS P#1165	152
CD11-DZMS P# 618	113	CD11-DZMS P# 1094	133	CD11-DZMS P#1169	153
P# 710	114	CD11-DZMS P# 1008	134	CD11-DZMS P#1206	154
CD11-DZMS P# 715	115	CD11-DZMS P# 1009	135	CD11-DZMS P#1231	155
CD11-DZMS P# 718	116	CD11-DZMS P# 1023	136	CD11-DZMS P#1233	156
CD11-DZMS P# 728	117	CD11-DZMS P# 1113	137	CD11-DZMS P#1237	157
CD11-DZMS P# 736	118	CD-DZMS P# 1134	138	CD11-DZMS P#1241	158
CD11-DZMS P# 738	119	CD11-DZMS P# 1135	139	CD11-DZMS P#1242	159
CD11-DZMS P#1245	160	CD11-DZMS P#1291	169	CD11-DZMS P#1363	178
Bakkalcha5	192	CD11-DZMS P#1304	170	CD11-DZMS P#1365	179
CD11-DZMS P#1248	161	Toltu	189	P#1367	180
CD11-DZMS P#1253	162	CD11-DZMS P#1316	171	Dire	190
CD11-DZMS P#1259	163	CD11-DZMS P#1318	172	P#1368	181
CD11-DZMS P#1260	164	CD11-DZMS P#1347	173	CD11-DZMS P#1369	182
CD11-DZMS P#1263	165	CD11-DZMS P#1353	174	CD11-DZMS P#1392	183
CD11-DZMS P#1268	166	CD11-DZMS P#1354	175	P#1445	184
CD11-DZMS P#1269	167	CD11-DZMS P#1355	176	CD11-DZMS P#1382	185
CD11-DZMS P#1290	168	CD11-DZMS P#1356	177	CD11-DZMS P#1411	186
		Ejersa	191	CD11-DZMS P#357	187
				CD11-DZMS P#719	188

Appendix B

Proc Mixed Codes Using SAS.

/* Model selection from candidate model*/;

/*MODEL1*/;

proc mixed data = ugbore22 method = REML covtest;

TITLE 'Augmented Block Design using PROC MIXED –genotypes (T.treat) is random and block and treatn are fixed';

Class block treat treatn; Model yield =block treatn; Random T.treat; **Run**;

 proc mixed data = ugbore22 method = ML covtest;

TITLE 'Augmented Block Design using PROC MIXED –genotypes (T.treat) is random and block and treatn are fixed';

Class block treat treatn; model yield =block treatn; random T.treat; **run**;

/*MODEL2*/;

proc mixed data = ugbore22 method = REML covtest; ;

TITLE 'Augmented Block Design using PROC MIXED – genotypes (T.treat) and block are random and treatn is fixed';

class block treat treatn; model yield = treatn ; random block T.treat; **run**;

proc mixed data = ugbore22 method = ML covtest; ;

TITLE ABD using PROC MIXED – genotypes (T.treat) and block are random and treatn is fixed'; class block treat treatn; model yield = treatn ; random block T.treat; **run**;

/*Check the residuals for normality */;

title "Checking Residuals for Normality";

proc univariate data=outreg1 PLOT NORMAL;var rstudent; histogram / normal; qqplot / normal(mu=est sigma=est);**run**;**quit proc means** data = ugbore22; var yield ; **run**;

 /*this procedure prints the mean of over all of yield that has only one value*/;

proc sort data=ugbore22; by block; **proc means** data=ugbore22; var yield; by block; **run**;

 /*this is used to print out the mean of yield in each block with 5 values*/

proc sort data=ugbore22; by treatn; **proc means** data=ugbore22; var yield; by treatn; **run**;

/* this is the procedure of calculating a mean of treatments and the means of four checks in all blocks that prints five values*/;

proc sort data=ugbore22; by treat; **proc means** data=ugbore22; var yield; by treat; **run**;

/*this is the procedure of calculating the mean of treat genotypes*/

/***PROC GLM for ABDs with One-Way Blocking**

SAS PROC GLM to obtain only intrablock and intravariety analyses for an ABD, as follows:*/

PROC GLM;TITLE 'Augmented block Design using GLM – Test treatmentsrandom and block and treatn are fixed ';

CLASS block treat treatn new; MODEL yield =treatn block T.treat/solution;

RANDOM T.treat/test; **RUN**;

LSMEANS treatn;

/*Variance Components Estimation methods*/

proc mixed data = ugbore22 method =type1 covtest; /*Variance Components Estimation Procedure using Henderson's Method III for trials designed in ABD*/

TITLE ' ABD using PROC MIXED – genotypes (T.treat) is random and treatn and block are fixed'; class block treat treatn; model yield = block treatn; random T.treat/ type=VC; **run**;

/*Variance Components Estimation Procedure using ML for trials designed in ABD*/

proc mixed data = ugbore22 method =ml asycov covtest ;

TITLE ABD using PROC MIXED – genotypes (T.treat) is random and treatn and block are fixed'; class block treat treatn; model yield = block treatn; random   T.treat/ type=vc; **run**;

proc mixed data = ugbore22 method =reml asycov covtest; /*Variance Components Estimation Procedure using REML for trials designed in ABD*/;

TITLE ABD using PROC MIXED – genotypes (T.treat) is random and treatn and block are fixed'; class block treat treatn; model yield = block treatn; random   T.treat/ type=vc; **run**;

proc mixed data = ugbore22 method =reml asycov covtest ;

TITLE ' ABD using PROC MIXED – genotypes (T.treat) is random and treatn and block are fixed'; CLASS block treat treatn; MODEL yield=treatn block/solution;

RANDOM T.treat/solution; LSMEANS treatn/diff; **RUN**;

Appendix C

Proc Mixed Outputs Using SAS.

Table C1: Constant Outputs for All Procedure

Model Information		Dimensions	
Data Set	WORK.UGBORE22	Covariance Parameters	2
Dependent Variable	yield	Columns in X	11
Covariance Structure	Variance Components	Columns in Z	192
Estimation Method	ML	Subjects	1
Residual Variance Method	Profile	Max Obs Per Subject	208
Fixed Effects SE Method	Model-Based	Observations Used	208
Degrees of Freedom Method	Containment	Observations Not Used	0
		Total Observations	208

Class Level Information		
Class	Levels	Values
block	5	1 2 3 4 5
treat	192	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 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192
treatn	5	189 190 191 192 999

Table C2: Selected EBULP for random effect (T.treat).

Solution for Random Effects					Solution for Random Effects				
Effect	treat	EBLUP	Std Err Pred	Rank	Effect	treat	EBLUP	Std Err Pred	Rank
T.treat	165	61.543	32.9996	1	T.treat	95	26.5104	32.8992	40
T.treat	133	61.391	32.8992	2	T.treat	182	26.4008	32.9996	41
T.treat	136	58.606	32.8992	3	T.treat	164	25.5609	32.9996	42
T.treat	14	56.308	32.8992	4	T.treat	144	25.1434	32.8992	43
T.treat	76	50.453	32.8992	5	T.treat	188	24.9421	32.9996	44
T.treat	187	47.663	32.9996	6	T.treat	65	23.4883	32.8992	45
T.treat	55	47.226	32.8992	7	T.treat	129	23.331	32.8992	46
T.treat	106	44.944	32.8992	8	T.treat	75	21.6317	32.8992	47
T.treat	185	44.392	32.9996	9	T.treat	40	20.8124	32.8992	48
T.treat	19	42.207	32.8992	10	T.treat	29	20.5471	32.8992	49
T.treat	89	40.833	32.8992	11	T.treat	13	20.2819	32.8992	50
T.treat	115	40.523	32.8992	12	T.treat	73	19.2447	32.8992	51
T.treat	120	39.02	32.8992	13	T.treat	71	18.4932	32.8992	52
T.treat	51	37.943	32.8992	14	T.treat	21	18.3369	32.8992	53
T.treat	64	37.634	32.8992	15	T.treat	146	18.1149	32.8992	54
T.treat	184	37.364	32.9996	16	T.treat	105	17.8906	32.8992	55
T.treat	163	36.745	32.9996	17	T.treat	124	17.6728	32.8992	56
T.treat	24	36.593	32.8992	18	T.treat	126	17.6728	32.8992	57
T.treat	11	36.24	32.8992	19	T.treat	145	16.9214	32.8992	58
T.treat	154	34.824	32.8992	20	T.treat	49	16.9019	32.8992	59
T.treat	61	34.672	32.8992	21	T.treat	155	15.4626	32.8992	60
T.treat	8	34.118	32.8992	22	T.treat	43	15.2663	32.8992	61
T.treat	28	33.985	32.8992	23	T.treat	158	14.6227	32.8992	62
T.treat	38	33.764	32.8992	24	T.treat	122	14.1365	32.8992	63
T.treat	108	33.495	32.8992	25	T.treat	166	14.0678	32.9996	64
T.treat	98	33.406	32.8992	26	T.treat	33	13.7839	32.8992	65
T.treat	52	33.036	32.8992	27	T.treat	142	13.7828	32.8992	66
T.treat	57	32.727	32.8992	28	T.treat	134	13.4292	32.8992	67
T.treat	12	32.659	32.8992	29	T.treat	10	12.6346	32.8992	68
T.treat	34	31.775	32.8992	30	T.treat	157	12.5893	32.8992	69
T.treat	103	31.461	32.8992	31	T.treat	159	12.5009	32.8992	70
T.treat	112	30.489	32.8992	32	T.treat	41	12.3046	32.8992	71
T.treat	20	30.449	32.8992	33	T.treat	68	12.2162	32.8992	72
T.treat	47	29.235	32.8992	34	T.treat	109	12.1882	32.8992	73
T.treat	22	28.99	32.8992	35	T.treat	48	11.8626	32.8992	74

Effect	treat	EBLUP	Std Err Pred	Rank	Effect	treat	EBLUP	Std Err Pred	Rank
T.treat	78	28.925	32.8992	36	T.treat	45	11.7299	32.8992	75
T.treat	31	27.708	32.8992	37	T.treat	99	10.5526	32.8992	76
T.treat	39	26.78	32.8992	38	T.treat	179	10.5315	32.9996	77
T.treat	91	26.51	32.8992	39	T.treat	125	9.8045	32.8992	78
T.treat	86	9.6244	32.8992	79	T.treat	81	4.9387	32.8992	91
T.treat	56	9.1219	32.8992	80	T.treat	135	4.7652	32.8992	92
T.treat	85	8.475	32.8992	81	T.treat	54	3.729	32.8992	93
T.treat	59	8.282	32.8992	82	T.treat	156	2.8202	32.8992	94
T.treat	104	8.0772	32.8992	83	T.treat	17	2.7328	32.8992	95
T.treat	36	7.7721	32.8992	84	T.treat	117	1.977	32.8992	96
T.treat	102	7.5468	32.8992	85	T.treat	88	1.8002	32.8992	97
T.treat	131	6.4449	32.8992	86	T.treat	60	1.2977	32.8992	98
T.treat	118	6.2648	32.8992	87	T.treat	96	0.7393	32.8992	99
T.treat	30	5.8713	32.8992	88	T.treat	183	0.1435	32.9996	100
T.treat	114	5.4249	32.8992	89	T.treat	18	0.08054	32.8992	101
T.treat	173	4.9617	32.9996	90					

Table C3: Empirical Best Unbiased Linear Predictions for random factors.

Solution for Random Effects					Solution for Random Effects				
Effect	treat	EBLUP	Std Err Pred	Rank	Effect	treat	EBLUP	Std Err Pred	Rank
T.treat	165	61.5432	32.9996	1	T.treat	95	26.5104	32.8992	40
T.treat	133	61.3909	32.8992	2	T.treat	182	26.4008	32.9996	41
T.treat	136	58.606	32.8992	3	T.treat	164	25.5609	32.9996	42
T.treat	14	56.3084	32.8992	4	T.treat	144	25.1434	32.8992	43
T.treat	76	50.4529	32.8992	5	T.treat	188	24.9421	32.9996	44
T.treat	187	47.6631	32.9996	6	T.treat	65	23.4883	32.8992	45
T.treat	55	47.226	32.8992	7	T.treat	129	23.331	32.8992	46
T.treat	106	44.9436	32.8992	8	T.treat	75	21.6317	32.8992	47
T.treat	185	44.392	32.9996	9	T.treat	40	20.8124	32.8992	48
T.treat	19	42.2072	32.8992	10	T.treat	29	20.5471	32.8992	49
T.treat	89	40.8326	32.8992	11	T.treat	13	20.2819	32.8992	50
T.treat	115	40.5232	32.8992	12	T.treat	73	19.2447	32.8992	51
T.treat	120	39.0202	32.8992	13	T.treat	71	18.4932	32.8992	52
T.treat	51	37.9431	32.8992	14	T.treat	21	18.3369	32.8992	53
T.treat	64	37.6337	32.8992	15	T.treat	146	18.1149	32.8992	54
T.treat	184	37.3635	32.9996	16	T.treat	105	17.8906	32.8992	55
T.treat	163	36.7446	32.9996	17	T.treat	124	17.6728	32.8992	56
T.treat	24	36.5933	32.8992	18	T.treat	126	17.6728	32.8992	57
T.treat	11	36.2397	32.8992	19	T.treat	145	16.9214	32.8992	58

T.treat	154	34.8241	32.8992	20	T.treat	49	16.9019	32.8992	59
T.treat	61	34.672	32.8992	21	T.treat	155	15.4626	32.8992	60
T.treat	8	34.1179	32.8992	22	T.treat	43	15.2663	32.8992	61
T.treat	28	33.9852	32.8992	23	T.treat	158	14.6227	32.8992	62
T.treat	38	33.7642	32.8992	24	T.treat	122	14.1365	32.8992	63
T.treat	108	33.4947	32.8992	25	T.treat	166	14.0678	32.9996	64
T.treat	98	33.4063	32.8992	26	T.treat	33	13.7839	32.8992	65
T.treat	52	33.0364	32.8992	27	T.treat	142	13.7828	32.8992	66
T.treat	57	32.727	32.8992	28	T.treat	134	13.4292	32.8992	67
T.treat	12	32.6591	32.8992	29	T.treat	10	12.6346	32.8992	68
T.treat	34	31.775	32.8992	30	T.treat	157	12.5893	32.8992	69
T.treat	103	31.4613	32.8992	31	T.treat	159	12.5009	32.8992	70
T.treat	112	30.4888	32.8992	32	T.treat	41	12.3046	32.8992	71
T.treat	20	30.4489	32.8992	33	T.treat	68	12.2162	32.8992	72
T.treat	47	29.2348	32.8992	34	T.treat	109	12.1882	32.8992	73
T.treat	22	28.9902	32.8992	35	T.treat	48	11.8626	32.8992	74
T.treat	78	28.9254	32.8992	36	T.treat	45	11.7299	32.8992	75
T.treat	31	27.7082	32.8992	37	T.treat	99	10.5526	32.8992	76
T.treat	39	26.7799	32.8992	38	T.treat	179	10.5315	32.9996	77
T.treat	91	26.5104	32.8992	39	T.treat	125	9.8045	32.8992	78
T.treat	86	9.6244	32.8992	79	T.treat	67	-11.212	32.8992	123
T.treat	56	9.1219	32.8992	80	T.treat	138	-11.237	32.8992	124
T.treat	85	8.475	32.8992	81	T.treat	25	-11.722	32.8992	125
T.treat	59	8.282	32.8992	82	T.treat	87	-12.08	32.8992	126
T.treat	104	8.0772	32.8992	83	T.treat	151	-12.209	32.8992	127
T.treat	36	7.7721	32.8992	84	T.treat	116	-12.743	32.8992	128
T.treat	102	7.5468	32.8992	85	T.treat	127	-12.828	32.8992	129
T.treat	131	6.4449	32.8992	86	T.treat	130	-13.977	32.8992	130
T.treat	118	6.2648	32.8992	87	T.treat	180	-14.002	32.9996	131
T.treat	30	5.8713	32.8992	88	T.treat	170	-14.179	32.9996	132
T.treat	114	5.4249	32.8992	89	T.treat	107	-14.202	32.8992	133
T.treat	173	4.9617	32.9996	90	T.treat	101	-15.174	32.8992	134
T.treat	81	4.9387	32.8992	91	T.treat	53	-15.235	32.8992	135
T.treat	135	4.7652	32.8992	92	T.treat	123	-15.259	32.8992	136
T.treat	54	3.729	32.8992	93	T.treat	62	-16.428	32.8992	137
T.treat	156	2.8202	32.8992	94	T.treat	167	-17.008	32.9996	138
T.treat	17	2.7328	32.8992	95	T.treat	175	-17.096	32.9996	139
T.treat	117	1.977	32.8992	96	T.treat	74	-18.241	32.8992	140
T.treat	88	1.8002	32.8992	97	T.treat	178	-18.732	32.9996	141
T.treat	60	1.2977	32.8992	98	T.treat	69	-18.771	32.8992	142
T.treat	96	0.7393	32.8992	99	T.treat	141	-19.945	32.8992	143
T.treat	183	0.1435	32.9996	100	T.treat	1	-22.022	32.8992	144
T.treat	18	0.08054	32.8992	101	T.treat	50	-23.191	32.8992	145

T.treat	189	0	43.6655	102	T.treat	113	-24.369	32.8992	146
T.treat	190	0	43.6655	103	T.treat	121	-25.073	32.8992	147
T.treat	191	0	43.6655	104	T.treat	77	-25.269	32.8992	148
T.treat	192	0	43.6655	105	T.treat	174	-25.628	32.9996	149
T.treat	139	-1.5118	32.8992	106	T.treat	160	-25.692	32.8992	150
T.treat	181	-2.0225	32.9996	107	T.treat	93	-26.27	32.8992	151
T.treat	37	-2.1297	32.8992	108	T.treat	58	-26.772	32.8992	152
T.treat	148	-2.7496	32.8992	109	T.treat	70	-27.17	32.8992	153
T.treat	4	-3.6768	32.8992	110	T.treat	119	-28.48	32.8992	154
T.treat	80	-3.7858	32.8992	111	T.treat	169	-29.562	32.9996	155
T.treat	32	-4.5609	32.8992	112	T.treat	171	-29.606	32.9996	156
T.treat	27	-5.0472	32.8992	113	T.treat	83	-29.85	32.8992	157
T.treat	150	-5.1366	32.8992	114	T.treat	147	-29.935	32.8992	158
T.treat	186	-5.2495	32.9996	115	T.treat	42	-30.087	32.8992	159
T.treat	26	-6.8595	32.8992	116	T.treat	90	-30.292	32.8992	160
T.treat	132	-8.8055	32.8992	117	T.treat	162	-30.402	32.9996	161
T.treat	82	-8.8972	32.8992	118	T.treat	16	-30.73	32.8992	162
T.treat	153	-8.9824	32.8992	119	T.treat	137	-32.411	32.8992	163
T.treat	94	-8.9857	32.8992	120	T.treat	63	-32.784	32.8992	164
T.treat	111	-9.7371	32.8992	121	T.treat	172	-33.363	32.9996	165
T.treat	5	-10.086	32.8992	122	T.treat	46	-34.198	32.8992	166
T.treat	177	-34.512	32.9996	167	T.treat	143	-46.423	32.8992	182
T.treat	9	-35.459	32.8992	168	T.treat	15	-48.323	32.8992	183
T.treat	79	-35.657	32.8992	169	T.treat	168	-48.791	32.9996	184
T.treat	92	-35.773	32.8992	170	T.treat	161	-48.923	32.9996	185
T.treat	97	-35.861	32.8992	172	T.treat	152	-49.695	32.8992	186
T.treat	44	-37.602	32.8992	173	T.treat	84	-51.687	32.8992	187
T.treat	2	-37.979	32.8992	174	T.treat	128	-51.816	32.8992	188
T.treat	176	-38.049	32.9996	175	T.treat	23	-52.39	32.8992	189
T.treat	35	-40.366	32.8992	176	T.treat	100	-53.146	32.8992	190
T.treat	72	-40.386	32.8992	177	T.treat	66	-60.854	32.8992	191
T.treat	149	-43.550	32.8992	178	T.treat	6	-63.927	32.8992	192
T.treat	140	-44.080	32.8992	179	T.treat	3	-72.194	32.8992	193
T.treat	110	-45.852	32.8992	180	T.treat	84	-51.687	32.8992	187
T.treat	7	-45.936	32.8992	181	T.treat	128	-51.816	32.8992	188
T.treat	143	-46.423	32.8992	182	T.treat	23	-52.39	32.8992	189
T.treat	15	-48.323	32.8992	183	T.treat	100	-53.146	32.8992	190
T.treat	168	-48.790	32.9996	184	T.treat	66	-60.854	32.8992	191
T.treat	161	-48.923	32.9996	185	T.treat	6	-63.927	32.8992	192
T.treat	152	-49.694	32.8992	186	T.treat	3	-72.194	32.8992	193