आई.ए.एस.आर.आई / पी.आर–05/2021
IASRI/PR-05/2021
Institute Project Code: AGEDIASRISIL201801300122

वंशानुगत के रोबस्ट आंकलन पर अध्ययन

Study of Robust Estimation of Heritability

ऐ.के.पॉल, हिमाद्री सेखर रॉय, रंजित पॉल, एल.एम.भर (30 जुलाई , 2021 तक)

A. K.Paul , Himadri Sekhar Roy, Ranjit Paul, L.M.Bhar (up to 30th July,2021)

सांख्यिकीय आनुवंषिकी
Division of Statistical Genetics

भा.कृ.अनु.प.- भारतीय कृषि सांख्यिकी अनुसंधान संस्थान
लाइब्रेरी एवेन्यु, पूसा, नई दिल्ली –110 012
I.C.A.R.- Indian Agricultural Statistics Research Institute
Library Avenue, Pusa, New Delhi – 110 012
www.iasri. icar.gov.in

(2021)

आमुख

वंशानुगत एक महत्वपूर्ण आनुवंशिक पैरामीटर है और आनुवंशिकविद तथा प्रजनक के द्वारा मात्रात्मक ट्रेट में सुधार तथा चयन के लिये व्यापक रूप से प्रयोग किया जाता है। वंशानुगत का सही और सटीक ज्ञान एक सफल प्रजनन कार्यक्रम के लिये महत्वपूर्ण है। पौधे और पशु प्रजनन आंकड़ों में कई व्यवहारिक स्थितियों में देखा गया है कि प्रेक्षण को स्वतन्त्र रूप से वितरित नहीं किया गया है तथा प्रेक्षणों के बीच किसी प्रकार के सहसंबंध मौजूद है। आंकड़ों के सहसंबंध प्रवृतियों की उपस्थिति में, प्रेक्षणों के बीच स्वतन्त्रता की क्लासीकल धारणा का उल्लंघन किया गया है। महत्वपूर्ण करेक्टर के प्रसरण के आनुवंशिक घटकों पर सूचना पौधे और पशु प्रजनकों में प्रथम रूचि हैं। इस प्रकार, विभिन्न आनुवंशिक प्रसरणों तथा इनफेरिंग, के आंकलन के बारें में उनके वंशानुगत हैं। विभिन्न आनुवंशिक पैरामीटर के आंकलन के आधार पर पौधें और पशु प्रजनक कार्यक्रम के दृष्टिकोण से बहुत महत्वपूर्ण है।

इसके अलावा विभिन्न कारकों द्वारा परिवर्तनशीलता की व्याख्या की गई है। यह देखा गया है कि सहसंबंध प्रेक्षणों की उपस्थिति में, त्रुटि घटकों से कुछ अधिक संरचित किया जा सकता है। इसलिये वास्तविक यादृच्छिक त्रुटि घटक पर पहुँचा जा सके। इसलिये एक मॉडल देखने की आवश्यकता है जहाँ त्रुटि के कारण स्वतन्त्र या वास्तविक यादृच्छिक त्रुटि मॉडल पर पहुँचने के क्रम आगे संरचनात्मक बदलाव किया जा सकता है यह इसलिये सांख्यिकीय तकनीकों के विकास और प्रयोग के समर्थक जिसमें परिवर्तनशीलता की मात्रा विभिन्न कारणों की वजह दोनों आनुवंशिक एवं प्रारूपी स्तर पर वैज्ञानिक तरीके एवं कारक पर मूल्यांकन कर सकते हैं तथा प्रजनन मानों की उच्च मात्रा की परिशुद्धता तथा पूर्वानुमान की तुलना अधिक कुशलता से की सकती है। इस प्रकार आनुवंशिक पैरामीटर के आंकलन के लिये सांख्यिकीय दृष्टिकोण को विकसित करना जब त्रुटियाँ सांख्यिकीय आनुवंशिकी के क्षेत्र में एक महत्वपूर्ण शोध करने योग्य क्षेत्र के साथ सहसंबंध है। क्षेत्रीय प्रोयोगिक आंकड़ों के दृष्टिकोण के क्लासीकल विश्लेषण में, आंकड़ों में उपस्थित सहसंबंध की उपेक्षा की जाती है। आंकड़ों के सहसंबंध प्रवृत्तियों के महत्वपूर्ण उपस्थिति में हम सहसंबंध प्रभाव की उपेक्षा नहीं कर सकते हैं। इस कारण इसलिये, इन स्थितियों तथा विश्लेषण की विधियों को देखना आवश्यक हो जाता है क्योंकि प्रसरण घटकों के आंकलन करने के लिये सिद्धांत असहसंबंध त्रुटियों के लिये केवल साहित्य में उपस्थित है। इसलिए अध्ययन वंशानुगत के रोबस्ट आंकलन पर वर्तमान जांच मार्च, 2018 में शुरू की गई थी। यह परियोजना सितंबर, 2021 तक तक सक्रिय थी। वंशानुगत के आंकलन पर मात्रात्मक करेक्टर के वंशानुगत तथा गैर–आनुवंशिक कारकों के प्रभाव पर सहसंबंध त्रुटि के प्रभाव और सहसंबंध त्रुटियों की उपस्थिति में आनुवंशिकता के आकलन पर बाहरी प्रभाव के प्रभाव का समालोचनात्मक परीक्षण करके प्रकाशन के लिए समर्पित है ।

सहसंबंध त्रुटियों की उपस्थिति में अन्वेषण आई.सी.ए.आर.–आई.ए.एस.आर.आई. के सांख्यिकीय आनुवंशिकी प्रभाग के अनुसंधान कार्यक्रम के एक भाग के रूप में किया गया था। यह जांच भाकृअनुप–भाकृसांअसं के सांख्यिकीय आनुवंशिकी प्रभाग के अनुसंधान कार्यक्रम के एक भाग के रूप में की गई थी। अध्ययन डॉ. एल.एम.भर, कार्यवाहक निदेशक और सांख्यिकीय आनुवंशिकी प्रभाग के प्रमुख, डॉ. तौकीर अहमद, कार्यवाहक निदेशक, डॉ. राजेंद्र प्रसाद, निदेशक भाकृअनुप–भाकृसांअसं के मार्गदर्शन में आयोजित किया गया था। समीक्षात्मक रूप से रिपोर्ट को पढ़ने और उसमें सुधार के लिए उपयोगी सुझाव देने के लिए लेखक रेफरी के बहुत आभारी हैं। श्रीमति सविता वधवा, मुख्य तकनीकी अधिकारी, श्री.नितिन जोशी, तकनीकी सहायक (टी–3), श्री. एस.पी.सिंह, सहायक मुख्य तकनीकी अधिकारी, विश्लेषण, सारणीकरण और रिपोर्ट तैयार करने में उनकी बहुमूल्य सहायता करने के लिये धन्यवाद करते हैं।

अंत में यह आशा की जाती है कि इस अध्ययन से प्राप्त परिणाम कुशल चयन और प्रजनन कार्यक्रम बनाने के लिए प्रजनकों के लिए उपयोगी होंगे।

सितंबर, 2021 लेखक

नई दिल्ली

PREFACE

Heritability is an important genetic parameter and is widely used by the geneticist and breeder for selection and improvement in quantitative traits. The precise and accurate knowledge of heritability is very important for the success a breeding programme. In the plant and animal breeding data it has been observed in many practical situations that observations are not independently distributed and some kind of correlation among the observations exists. In the presence of significantly correlated trends in the data, the classical assumption of independence between observations is violated. The information on genetic components of variances of the important characters is the prime interest of plant and animal breeders. Thus, the estimation of various genetic variances and inferring about their inheritance, based on estimates of the different genetic parameters is very important from plant and animal breeding programme point of view.

In addition to variability explained by different factors, it is seen that in the presence of correlated observations, something more from the error components can be structured so as to arrive at truly random error component. So, there is a need to look into a model wherein structural variation due to error could be further accounted for in order to arrive at independent or truly random error model. This, therefore, advocates the use and development of statistical techniques wherein amount of variability due to different causes both at genetic and phenotypic level can be assessed in a scientific manner and factors can be compared with fairly high amount of precision and prediction of the breeding values is carried out more efficiently.

Hence, developing statistical approach for estimation of genetic parameters when errors are correlated as well as presence of outlier, is an important researchable area in the arena of statistical genetics. In the classical analysis approach of the field experimental data, the correlation present in the data is ignored. In the presence of significantly correlated trends in the data, we cannot neglect the correlation effect. Because of this reason it therefore, becomes necessary to look for these situations and methods of analysis. Since theories for estimating the variance components available in the literature are only for uncorrelated errors, So the present investigation on Study of Robust Estimation of Heritability was initiated in March, 2018.The project was in operation till September, 2021. The publication is devoted to critically examine

the the influence of correlated errors on the inheritance of quantitative characters and the effect of outlier on the estimation of heritability in the presence of correlated errors.

The investigation was undertaken as a part of the research program of Statistical Genetics Division of ICAR-IASRI. The study was conducted under the guidance of Dr. L.M.Bhar, Acting Director and Head of Statistical Genetics Division, Dr. Tauqueer Ahamed, Acting Director, Dr. Rajendra Parsad, Director ICAR-IASRI. The authors are extremely thankful the refrees for critically going through the report and making useful suggessions to improve the same. Our thanks are also due Mrs SavitaWadhwa technical officer (T9) , Sh. Nitin Joshi, technical officer (T-3) , Sh. S.P.Singh, Technical officer(T-7-8) , for their valuable help in the analysis, tabulation and preparation of the report.

In the end it is hoped that results emerging from this study will be useful for breeders for making efficient selection and breeding program.

September,2021 AUTHORS

New Delhi

Content

INTRODUCTION

1. Background

Occurrence of outlier(s) is very common in every field in which data collection is involved. An outlier in a set of data is an observation (or an observation vector) that appears to be inconsistent with the remainder of the observations in that data set. Outlier(s) in genetics and breeding experiments is/ are also likely to appear. The impact of outlier in plant breeding is erroneous results in heritability estimation and genomic prediction. Here, our objective is to develop a suitable robust procedure for heritability estimation. The existing methods are not robust against the presence of outlier and non-normality of error or random effect for estimating heritability. There is a need to modify the existing methods or to develop the robust procedure for estimating heritability under different distributional assumptions of the random effect as well as the error term.

Since in most of the practical situations in plant as well as animal breeding data the assumption of independent errors is not fulfilled and they may be correlated. There is little work (no work) towards the development of methods for estimation of heritability in case of correlated error and correlated random effect for common dam as well as common dam and sire. So there is need to develop methodology for estimation of important genetic parameter heritability by considering the correlated random effects and error structure in the models.

The assumptions of uncorrelated error in Henderson's mixed model do not hold always in practical situation. Therefore, there is a need to modify the above model for correlated error as well as random effect. The researchers estimate genetic parameters by assuming independent errors. The consequence of this is the upward or downward biased estimates of genetic parameters depending upon the nature of correlation among errors. So there is a need to develop breeding value estimation procedure using Henderson's Mixed model under correlated errors.

In the plant and animal breeding data it has been observed in many practical situations that observations are not independently distributed and some kind of correlation among the observations existscorrelation (Falconer, 1989). In the presence of significantly correlated trends in the data, the classical assumption of independence between observations is violated. The information on genetic components of variances of the important characters is the prime interest of plant and animal breeders. Thus, the estimation of various genetic variances and inferring about their inheritance, based on estimates of the different genetic parameters is

very important from plant and animal breeding programme point of viewcorrelation (Falconer, 1989). In addition to variability explained by different factors, it is seen that in the presence of correlated observations, something more from the error components can be structured so as to arrive at truly random error component. So, there is a need to look into a model wherein structural variation due to error could be further accounted for in order to arrive at independent or truly random error model. This, therefore, advocates the use and development of statistical techniques wherein amount of variability due to different causes both at genetic and phenotypic level can be assessed in a scientific manner and factors can be compared with fairly high amount of precision and prediction of the breeding values is carried out more efficiently.

In literature, a large number of statistical techniques are available for analysis of data for random and fixed effects. The development of methods of estimating variance components was initiated a long age in the early 20^{th} century. Fisher (1925) made a major contribution to variance component models through initiating the concept of the analysis of variance method of estimation. Jackson (1939) dealt with a mixed model for the first time in the literature of variance component estimation. Cochran (1939) initiated for unbalanced data. Henderson (1953) gave a method of how to use unbalanced data for estimating variance components in a difficult situation. In the light of the weaknesses (negativity, lack of distributional properties) of ANOVA (Analysis of Variance) estimators developed earlier, other approaches namely ML(Maximum Likelihood) REML(Restricted Maximum Likelihood) , MINQUE(Minimum Norm Quadratic) etc. were emerged. Again, for examining the relationship among observations, different procedures are available in the literature. Durbin and Watson (1950) gave a procedure to check the presence of first order autocorrelation disturbance in the error term. Diblasi and Bowman (2001) proposed a test statistic and a graphical method, which can assess the evidence for the presence of any spatial correlation in the data. Wahi and Rao (2004) studied the effect of fixed effects on the estimate of heritability.Singh*et al* (2006) estimate variance components when errors are correlated in half sib data. Costa *et al* (2009) use autoregressive multiple lactation animal model for estimating genetic parameters of test day fat and protein yields. Hence, developing statistical approach for estimation of genetic parameters when errors are correlated is an important researchable area in the arena of statistical genetics. In the classical analysis approach of the field experimental data, the correlation present in the data is ignored. In the presence of significantly correlated trends in the data, we cannot neglect the correlation effect. Because of this reason it therefore, becomes necessary to look for these situations and methods of analysis. Since theories for estimating

the variance components available in the literature are only for uncorrelated errors, so the present investigation were taken with the following objectives:

2. Objectives:

i. Development of robust procedure for heritability estimation
ii. Development of heritability estimation procedure in case of correlated random effect as well as error
iii. Development of breeding value estimation procedure using Henderson's Mixed model under correlated errors
iv. Development of appropriate computer programme for the developed methodologies using SAS/R.

1.1 Introduction:

The linear mixed model (LMM) is employed in the evaluation of the genetic association of continuous traits to examine the association between the phenotype and molecular marker. LMM also serve as the foundation for estimating SNP-based heritability, which is calculated using the LMM estimated variance. The LMM is based on a number of key assumptions (Laird and Ware, 1982), and caution should be exercised when data contamination (outliers) leads to a violation of the assumption of the normality of the errors, because not only association analysis results, but also variance component estimation and the coefficient of determination can be skewed (Lourenc *et al.*, 2011, Demidenko, 2013; Rousseeuw, 1984). In general, it is well understood that the existence of even a single outlier can have a negative impact on methods based on likelihood (Huber, 1964). Real-world experiments generate data with unusual observations that are not always due to measurement errors. Outliers in plant breeding experiments might result from inherent properties of the genotypes investigated, habitats, or even years (Bernal-Vasquez *et al.*, 2016; Estaghvirou *et al.*, 2014). As a result, it is critical to not only check the assumptions of the model, but also develop alternative methodologies capable of producing accurate findings when the data deviates from the model's main assumptions. To minimize the effect of outlying observations, robust procedures are designed. A goodness-of-fit measure of interest that can be seriously affected by data contamination in the classical coefficient of determination (Croux and Dehon, 2003; Rousseeuw, 1984). Data contamination can affect the classical coefficient of determination, which is an essential indication of goodness-of-fit.

Robust approaches have been given for dealing with normality violations and other model misspecifications in the simple linear regression model, and we may identify various robust alternatives to the classical coefficient of determination in the context of robust linear regression (Maronna *et al.*, 2006; Renaud and Vitoria-Feser, 2010). When models are fitted using restricted maximum likelihood, however, the likelihoods of LMMs with different fixed effects structures cannot always be compared (REML; Pinheiro and Bates, 2000). Maximum likelihood (ML) estimation solves the problem, but the variance component estimations will be more biased as a result (Pinheiro and Bates, 2000).In animal breeding, derivative-free (DF) techniques are most

commonly used in variance component estimation by REML (Graser *et al.*, 1987). The fundamental reason for this is that they just require the REML log-likelihood function to be computed, as well as a maximization procedure to determine its maximum. In this method, the direct inversion of the mixed model equations' coefficient matrix, whose size can equal the whole number of levels of all fitted random effects, can be readily avoided. Although DF-REML methods have been shown to be the most suitable for univariate analyses (Meyer, 1994), they have also been shown to be computationally expensive and/or inaccurate in multi-trait analyses, in which case second-derivative algorithms (e.g. Newton-Rapshon, modified Newton-Raphson EM, or others) are highly advantageous. In the case of our application, DF-REML is suitable, and it has the further advantage of being easily robustified by suitably plugging-in robust estimators in the optimization phase.

In this study, we have implemented robust heritability estimation approaches for estimation SNP-based heritability when observations in the dataset may be contaminated with outlying observations. The robust DF-REML estimate method have been described for estimating variance components using LMMs. Huber's (1964) and Rousseeuw and Croux's (1993) robust estimates of scale was implemented for estimation procedure and the computation of the coefficients of determination, respectively. A real dataset was applied using the robust methods and different procedures are followed such as heritability and variance components are estimated using the original data as well as contaminated data.

1.2 Robust estimation in the LMM

A general Linear mixed model can be written as-

$$Y=X\beta+Zu+e \quad (1)$$

Where $y_{n\times 1}$is a vector of observations,$X_{n\times(p+1)}$ is a design matrix for the fixed effects (intercept included), $\beta_{(p+1)\times 1}$ is the vector of unknown fixed effects, $z_{n\times q}$is the design matrix for the randomeffects, $u_{q\times 1}$~N(0,G)is the vector of unknown random effects and $e_{n\times 1}$~ N(0,R) is the vector of random errors. It is assumed that cov(u,e) = 0 and y~N(Xβ,ZGZ'+R). Here, G and R are unknown; R =$\sigma_e^2 I$, with σ_e^2 a residual variance which needs to be estimated, which means that the errors have equal variances and are mutually independent.

Setting R=$\sigma_e^2 I$ usually constitutes an approximation to the structure of the error variance but, in

the context of genomic selection in plant breeding, might have the advantage of accounting for possible unexplained genetic variance, which can be absorbed by the error variance (Piepho *et al.*, 2012).

Here it is considered that R = $\sigma_e^2 I$ and G = $\sigma_e^2 A(\theta)$, where the variance matrix of the random effects A depends on the vector of *l* unknown variance parameters = $\sigma_1^2 \dots\dots\dots, \sigma_1^2/\sigma_e^2$.

Following these assumptions, the variance expression of y may be shown as follows:

$$\text{Var(y)} = \sigma_e^2 ZA(\theta)Z' + \sigma_e^2 I = \sigma_e^2 \phi,$$

With $\phi = ZA(\Theta)Z' + I$

As we know that, in case of REML estimation procedure, prior to estimating the fixed effects components, the variance components are estimated. Considering transformation of model (1) through a matrix$U_{(n-(p+1))\times n}$such that UU'= I - P and U'U=I , considering P as the orthogonal projection matrix on X. So, U'Xβ = 0 and therefore model (1) can be written as follows:

$$y^{\circ} = U'y = U'(Zu+e) \quad (2)$$

where, var(y°) = var(U'y) = σ_e^2 U'ϕU = $\sigma_e^2 \Phi^{\circ}$ and hence $y^{\circ} \sim N(0, \sigma_e^2 \Phi^{\circ})$

The REML log-likelihood is written as-

$$-2l(\sigma_e^2,\theta,y) \sim (n-(p+1))\log\sigma_e^2 + \log\left|U'\phi U\right| + 1/\sigma_e^2 (y'\phi^{-1}y) \quad (3)$$

If the Φ matrix is known (and thus Φ° is known), that means the variances $\sigma_1^2 \dots\dots\dots, \sigma_1^2$ are are also known, hence the estimate of the residual variance σ_e^2 can be expressed as-

$$\hat{\sigma}_e^2 = \frac{1}{n-(p+1)} y'\phi^{-1}y. \quad (4)$$

The REML profile-log-likelihood is expressed after replacing the model expression in (3)-

$$p_{lik}\left(\theta\middle|\hat{\sigma}_e^2, y\right) = -2l_p\left(\theta\middle|\hat{\sigma}_e^2, y\right) \sim (n-(p+1))\log\hat{\sigma}_e^2 + \log\left|\phi\right| \quad .$$

(5)

Inversion of the variance-covariance matrix, which is frequently non-diagonal and high dimensional, is required in Formulation (4).Its size is determined by the number of individuals involved, which can range from a few hundreds to thousands in genetic breeding research (Mrode, 2000; Searle, 1971). Due to numerical instability, direct inversion should be avoided if at all possible, and the use of a matrix decomposition, such as the Cholesky decomposition, is

recommended (Czkova and Czek, 2012). Φ°

If Φ is a symmetric positive-definite matrix (which implies that Φ° is also symmetric positive-definite), one can take the inverse of the Cholesky decomposition of Φ°; = $\Delta = (\text{chol}(\Phi^{\circ}))^{-1}$, and transform model (2) to

$$y^{*} = e^{*} \quad (6)$$

Now let $y^{*} = \Delta y^{\circ}$ and $e^{*} = \Delta U'(Zu + e)$. Hence var(y^{*}) = $\sigma_e^2 I$ and E(y^{*}) = 0, it immediately follows that $y^{*} \sim N(0, \sigma_e^2 I)$

It can be demonstrated that the log-likelihood of (6) has the same formulation as the log-likelihood of (3). Furthermore, the estimate of the residual variance is shown as-

$$\hat{\sigma}_e^2 = \frac{1}{n-(p+1)} (y^{*})' y^{*}, \quad (7)$$

The above model is equivalent to (4). This estimator of scale is applied in DF-REML procedure provide an initial estimate of scale in the optimization process (in this case the minimization of the profile-log-likelihood (5) and where θ is also estimated (Graser *et al*., 1987; Meyer,1989).

The robust DF-REML, initial robust estimates of scale need to be considered, e.g. the Huber's proposal-II (Huber, 1964), the robust location-free scale estimate of Rousseeuw and Croux (1993) or other robust estimates of scale. Only the two previous robust scale estimators are considered here.

after Robust estimation of θ has been done , the robust estimated $\hat{\Phi}^{0}$ and $\hat{\Delta}$ matrices and subsequently robust estimate of σ_e^2 and $\tilde{\sigma}_e^2$ were recalculated. The robust estimates of the random effects are obtained as $(\sigma_1^2, \ldots\ldots, \sigma_l^2) = \hat{\sigma}_e^2 \times \hat{\theta}$. Next, taking $\Delta = (\text{chol}(\Phi^{*}))^{-1}$, M-regression is applied to robustly estimate the vector of fixed effects β via the fit of the model

$$\Delta^{\bullet} y = \Delta^{\bullet} X\beta + e \quad (8)$$

where $e = \Delta^{\bullet}(Zu + e) \sim N(0, \sigma_e^2 I)$

and the distribution of the observation follows $\Delta^{\bullet} y \sim N(X\beta, \sigma_e^2 I)$

Details on the robust estimation of parameters is given below as a flowchart. The algorithms for implementing classical and Robust DF-REML estimation algorithm procedure is shown as

follows:

Table 1: Step by step procedure for Robust DF-REML

Step	**Model**	**Expected mean**	**Variance**	**Distribution**
1	Y=Xβ+Iu+e	$E(y) = X\beta$	$V(y) = 2\sigma^2 K + \sigma_e^2 I$ $= \sigma_e^2 (2\sigma_*^2 K + I)$ $= \sigma_e^2 \Phi$	$N(X\beta, \sigma_e^2 \Phi)$
Consider a transformation of step-1 model through an $n \times (n-r)$ matrix $U : U^T X\beta = 0$				
2	$y = U^T y = (X\beta + Iu + e)$ $= U^T (Iu + e)$	$E(y^\circ) = 0$	$V(y^\circ) = \sigma_e^2 U^T \Phi U$ $= \sigma_e^2 \Phi^\circ$	$N(0, \sigma_e^2 \Phi^\circ)$
Assuming Φ° is known, step-2 model is considered and transform as $\Delta = (\text{chol}(\Phi^\circ))^{-1}$				
3	$y^* = \Delta y^\circ = \Delta U^T (Iu + e)$	$E(y^*) = 0$	$V(y^*) = \sigma_e^2 I$	$N(0, \sigma_e^2 I)$
4	Now we need to estimate σ_e^2 from step-3 model Classical REML: $\hat{\sigma}_e^2 = y^{*T} y^* / (n - (p+1))$ Initial estimate of scale Robust REML: $\hat{\sigma}_e^2 = Q_n(y^*)^2$ $\text{MAD}(y^*)^2$, $\text{huber's}(y^*)^2$ or others			
5	Get the profile log-likelihood from step-2: $I_p(\sigma_*^2 \mid \sigma_e^2 = \hat{\sigma}_e^2, y) = c - \frac{n-(p+1)}{2} \log(\hat{\sigma}_e^2) - \frac{1}{2} \log \lvert \Phi^\circ \rvert$			
6	Plug-in the classical and robust estimates $\hat{\sigma}_e^2$ obtained in step 4 in the profile log-likelihood function l_P and optimize for σ_*^2			
7	Now we need to compute $\hat{\Phi}^\circ = U^T (2\sigma_*^2 K + I) U, \hat{\Delta} = (\text{chol}(\hat{\Phi}))^{-1}$ and final $\hat{\sigma}_e^2$ from step-3 model as : Classical REML: $\hat{\sigma}_e^2 = \text{var}(\hat{\Delta} y^\circ)$ Robust REML: $\hat{\sigma}_e^2 = Q_n(\hat{\Delta} y^\circ)^2$ $\text{MAD}(\hat{\Delta} y^\circ)^2$, $\text{huber's}(\hat{\Delta} y^\circ)^2$ or others			
8	Calculate $\sigma_e^2 = \sigma_*^2 \times \hat{\sigma}_e^2$ for both classical and robust estimation			
9	Consider a transformation of model in step-1 through	$E(y^\bullet) = X^\bullet \beta$	$V(y^\bullet) = \sigma_e^2 I$	$N(X^\bullet \beta, \sigma_e^2 I)$

	matrix $\Delta^{\bullet} = (\text{chol}(\hat{\Phi}))^{-1}$ $y^{\bullet} = \Delta^{\bullet} y = \Delta^{\bullet}(X\beta + Iu + e)$ $= X^{\bullet}\beta + e$			
10	Estimate the fixed parameters β via model (4) classical REML: classical least-squares estimation robust REML: M-regression			

1.3 Estimation of robust coefficient of determination:

Nakagawa and Schielzeth (2013) proposed a general method for obtaining R^2 from generalized linear mixed-effects models. They defined a coefficient of determination for the variance explained by the fixed factors, marginal R_m^2, and another for the variance explained by both the fixed and random factors, conditional R_c^2.

A general technique for determining R^2 from generalized linear mixed-effects models was proposed by Nakagawa and Schielzeth (2013). They explained a marginal coefficient of determination for the variance explained by fixed factors, marginal $R_m^2$and another by using both the fixed and random factors and conditional R_c^2.Those coefficients rewrite in the LMM as

$$R_m^2 = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_e^2 + \sigma^2} \text{ and } R_m^2 = \frac{\sigma_f^2 + \sigma^2}{\sigma_f^2 + \sigma_e^2 + \sigma^2} \quad (9)$$

Here σ_e^2 and σ^2 are residual variance and the unknown genetic variance respectively. $\sigma_f^2$is the variance of fixed effect components. σ_f^2can be estimated by the sample variance of the elements of the vector $X\hat{\beta}$, i.e. $\hat{\sigma}_f^2 = var(X\hat{\beta})$, here degrees-of- freedom correction (Snijders and Bosker, 1999) can be avoided.

Here, for the robust DF-REML approach, esti mate σ_f^2 is done by using the robust location-free scale estimator Q_n of Rousseeuw and Croux (1993), the Huber M-estimator of scale (Huber, 1981). Both the estimators are implemented using R packages MASS and robust base, under the functions Hubers() and Qn (), respectively. These functions are used to estimateσ_f^2 and $\hat{\sigma}_f^2 = Q_n(X\hat{\beta})^2$and $\hat{\sigma}_f^2 = (hubers(X\hat{\beta})\$s)^2$, respectively.

1.4 Linear Mixed Models for heritability estimation:

In quantitative genetics, the linear mixed effects model proposed by Yu *et al.* (2006), which assumes a single observation per genotype and is shown as-

$$Y=X\beta+u+e \quad (10)$$

where y is the vector of phenotypic observations, σ^2 is an unknown genetic variance; and e ~ N(0,R = $\sigma_e^2 D$) is the vector of random errors, with σ_e^2 an unknown residual variance, and D a diagonal matrix with entries equal to $1/r_i$, where r_i equals to the number of observations from which each phenotypic data point was obatined. When there is no information on D then D = I is considered. In any case, rescaling the data with the weight matrix $D^* = D^{-1/2}$ allocates us to the case where R = σ_e^2I and therefore we herein consider, without loss of generality, that D = I. Hence, we have that y ~N(Xβ, $2\sigma^2$K+ σ_e^2I)with K a kinship matrix. Since var(y) = $\sigma_e^2\left(\frac{2K\sigma^2}{\sigma_e^2} + I\right) = \sigma_e^2(2K\sigma^2 + I) = V$, the vector ofunknown variance components that needs to be estimated is the two dimensional vector θ = (σ_*^2, σ_e^2).

Estimation of narrow-sense heritability is crucial in plant and animal breeding for determining the response to selection and in human studies for evaluating the potential to predict genetic risk for disease, among other things. The general formula for estimating heritability is given as

$$h^2 = \sigma^2/(\sigma^2 + \sigma_e^2) \quad (11)$$

where σ^2 and σ_e^2 are the genetic additive and residual variances. Note that (i) only a single observation per genotype is considered; (ii) u ~N(0, σ^2G) with G a genetic/GRM; and (iii) e ~ iid N(0, σ_e^2 I).

Point wise SNP-heritability (h^2_{SNP})proposed by Yang *et al.*(2014) suggest to estimate h^2_{SNP} as

$$h^2_{SNP} = (\hat{\sigma}^2 + \hat{\sigma}_e^2 - \hat{\hat{\sigma}}^2 - \hat{\hat{\sigma}}^2_e)/(\hat{\sigma}^2 + \hat{\sigma}_e^2) \quad (12)$$

where $\hat{\sigma}^2$ and $\hat{\sigma}_e^2$ are the estimates obtained by the fit of model (10) with SNP removed from the fixed effects component, and $\hat{\hat{\sigma}}^2$ and $\hat{\hat{\sigma}}^2_e$ are the estimates obtained by the fit of model (10). In

this work propose, based on the previous model formulations, to estimate h^2_{SNP} as the difference between the marginal R^2 obtained from the fit of the full model (10) and the marginal R^2 obtained from the fit ofmodel (10) without the SNP (R^2 computed as in Section 3), i.e. as

$$h^2_{SNP} = R^2_{m;(10)} - R^2_{m;(10)SNP-out^*} \tag{15}$$

The robust DF-REML estimation approach is presented in table 1 as a step-by-step methodology for the estimation of variance components in LMMs with only one random effects variance component, such as models The acronyms DF-REML-h and DF-REML-Q_n are used to the robust DF-REML methods that use Huber's (1964) and Rousseeuw and Croux's (1993). Below we have briefly discussed M-regression, Huber's (1964) and Rousseeuw and Croux's (1993) respectively.

1.5 M-regression:

The general multiple linear regression model can be written as

$$\mathrm{y} = \mathrm{X}\beta + \mathrm{e} \tag{16}$$

under the usual assumptions of independence, homoscedasticity and normality of the errors,where $\mathrm{y}_{n\times 1}$ is a vector of observations, $\mathrm{X}_{n\times p}$ is the design matrix for the fixed effects (intercept included), $\beta_{p\times 1}$ is the vector of unknown fixed effects, and $\mathrm{e}_{n\times 1}$ is a vector of non-observable errors with E(e) = 0 and covariance matrix $\mathrm{var}(\mathrm{e}) = \sigma^2 \mathrm{I}_n$. Note that we are considering the case where n > p.

The least squares (LS) estimate of β is obtained by minimizing the residual sum of squares,

$$\beta_{\mathrm{LS}} = \arg\min_{\beta} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\mathrm{y_i} - \mathrm{X_{i\bullet}}\beta\right)^2 \tag{17}$$

where $\mathrm{X_{i\bullet}} = \left(1, \mathrm{X_{i1}}, \cdots, \mathrm{X_{ip}}\right)$. If X has rank $\mathrm{P} \le \mathrm{n}$ we obtain $\hat{\beta}_{\mathrm{LS}} = \left(\mathrm{X^T X}\right)^{-1} \mathrm{X^T y}$, with covariance matrix $\sum \beta_{\mathrm{LS}} = \sigma^2 \left(\mathrm{X^T X}\right)^{-1}$. Since in the classical approach we have $\mathrm{e} \sim \mathrm{N}\left(0, \sigma^2 \mathrm{I_n}\right)$ then $\hat{\beta}_{\mathrm{LS}}$ is also the maximum likelihood estimate (MLE) of β. Additionally $\hat{\beta}_{\mathrm{LS}}$ is an UMVUE estimator of β (Gauss-Markov theorem).

When considering robust approaches in the estimation of the unknown parameters, the normality condition on the error distribution is relaxed and the estimators are obtained by methods other than maximum likelihood. In this study we restrict our attention to M-estimators, which have been shown to have good computational and efficiency properties(Maronna *et al.*, 2006; Rousseeuw& Leroy, 2003). In the M-regression approach, the β estimates are the solutions to the following minimization problem,

$$\hat{\beta}_R = \arg\min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{r_i(\beta)}{\hat{\sigma}}\right) \tag{18}$$

where ρ is an appropriate function, $r_i(\beta) = y_i - X_{i\bullet}\beta$ and $\hat{\sigma}$ some robust estimate of scale,usually the re-scaled mean absolute deviation (MADn; Rousseeuw and Croux, 1993):

$$MAD_n(X) = 1.483 \times med_i\left\{x_i - med_j\left(x_j\right)\right\} \tag{19}$$

One can easily see that $\rho(x) = x^2$ then $\hat{\beta}_R$ is the least squares estimate. Differentiating (3) with respect to every β_j , and equating to zero we get the p score equations

$$\sum_{i=1}^{n} \psi\left(\frac{r_i\left(\beta_j\right)}{\hat{\sigma}}\right) X_{ij} = 0, \qquad j = 1, \cdots, p \tag{20}$$

Although (3) and (5) are not always equivalent, (5) is useful in the search for solutions to (3). Moreover considering the weights

$$w_i = \psi\left(\frac{r_i\left(\beta_j\right)}{\hat{\sigma}}\right) \Big/ r_i\left(\beta_j\right) \qquad j = 1, \cdots, p \tag{21}$$

leads to $\hat{\beta}_R = \left(X^T W X\right)^{-1} X^T W y$, where W is a diagonal matrix with elements w_i. This shows that (5) can be solved by Iteratively Reweighted Least Squares (IRWLS) (Holland & Welsh, 1977; Maronna *et al*., 2006).

Within the class of M-estimators we have considered the _ function proposed by Huber(1964)

$$\rho(x) = \begin{cases} x^2/2 & \text{if} \quad |x| \le b \\ b\left(|x| - b/2\right) & \text{if} \quad |x| \rangle b \end{cases} \tag{22}$$

which assumes a quadratic form in the middle of the distribution and a linear one in the tails. Also, we have taken the tuning parameter b = 1:345 to achieve a 95% asymptotic relative efficiency when the data are normal and to allow for substantial resistance to outliers when the

data departs from the normality premise. Changing the value of b allows for different levels of compromise between efficiency and robustness. Specially, the bigger the value of b, the bigger the efficiency but the lesser the robustness.

It can easily be seen that in this case, the derivative of ρ writes as

$$\psi_b(x) = \begin{cases} x & |x| \le b \\ b & x > b \\ -b & \text{if } x < -b \end{cases} = \min(b, \max(-b, x)) \tag{23}$$

1.6 Estimates of scale for the robust DF-REML method

In the case of the proposed robust DF-REML approach, we consider two distinct estimators of scale:

(i) Huber's proposal II estimator (Huber, 1964; equation (9)), which, when consideringa tuning constant b = 1:5, is known to tolerate up to 26% contamination, i.e, has a26% break down point, and has 96% efficiency at Gaussian distributions. Huber'sproposal II estimator is defined by a function with two components,

$$\psi\left(z = \left(\psi_b(z), \psi_b(z)^2 - \beta\right)\right)$$

with ψ_b as in equation (8), $z = \frac{x - \mu}{\sigma}$ and $\beta = 2\Phi(b) - 1 - 2b\varphi(b) + 2b^2(1 - \Phi(b))$ with φ the density and distribution functions of the Gaussian distribution, respectively. Specially, the estimator is given as the solution to the following system of equations:

$$\sum_{i=1}^{n} \psi_b\left(\frac{x_i - \mu}{\sigma}\right) = 0 \text{ and } \sum_{i=1}^{n} \psi_b^2\left(\frac{x_i - \mu}{\sigma}\right) - \beta = 0$$

where the scale and location estimates need to be obtained by iterative procedures. This estimator corresponds to a combination of minimax results from the location and scale models, and the estimates of scale and location need to be estimated via iterative procedures (technical details can be found in Hampel *et al*., 2011).

(ii) The robust estimator of scale Qn (Rousseeuw and Croux, 1993; equation (10)), which presents a break down point point of 50% (the best possible one) but has only 82%efficiency at Gaussian distributions.

$$Q_n(x) = 2.219\left\{|x_i - x_j|; i<j\right\}_{(k)}$$

with $k \approx \binom{n}{2}/4$ the k-th order statistic of the $\binom{n}{2}$ inter-point distances.

Again, changing the values of the tuning constant b will allow for different compromises between Gaussian efficiency and robustness in the case of (i). More detail on this topic can be found in the works of, e.g., Huber (1964, 1981); Rousseeuw (1984); Rousseeuw and Croux (1993) and Rousseeuw& Leroy (2003).

1.7 Data description: A dataset has been collected from International Maize and Wheat Improvement Center (CIMMYT) of wheat improvement. The study consisted of 453 advanced parental lines from the CIMMYT hybrid wheat breeding program. The lines were planted in 2-m-long, double-row linear plots, with 20-cm inter-row spacing at El Batan, Mexico (20.83° N, 100.83° W) in 2017 and 2019, and at Obregon, Mexico (27.48° N, 109.93° W) in 2018 growing cycles. The population was genotyped with the 20K Infinium is elect SNP array by Trait Genetics (Gatersleben, Germany). The marker dataset was filtered for polymorphism and minor allele frequency (<5%) and >50% missing data were removed.

1.8 Result and discussion:

We have applied DF-REML estimates in LMMs to to incorporate Huber's proposal II estimator and robust estimator of scale Qn (Rousseeuw and Croux, 1993). Both the estimators are well known for their property of robustness. A real dataset (Wheat dataset) was analysed to detect the impact of the both robust estimators. Finally these estimators are compared with classical Linear mixed model and it is seen that robust estimators performed well as compared to the classical model. Table 2 describes the estimated variance components of the three different approaches and table 4 describes estimated heritability of three different methods. Here we have incorporated 1%, 3% and 5% contamination and checked that incorporation of outlying observations changes the estimated heritability values.

Table 2: Estimated variance component

	Without contaminated		1% contaminated		5% contaminated		10% contaminated	
	$\hat{\sigma}^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}^2$	$\hat{\sigma}_e^2$	$\hat{\sigma}^2$	$\hat{\sigma}_e^2$
Classical	0.41	0.73	0.39	0.86	0.39	1.01	0.43	1.30
Huber	0.33	0.58	0.33	0.65	0.38	0.74	0.37	0.80
Q_n	0.35	0.62	0.33	0.64	0.36	0.73	0.34	0.76

Table 4: Comparison of estimated heritability using different methods

	Heritability			
	Without contaminated	1% contaminated	5% contaminated	10% contaminated
Classical	0.36	0.31	0.28	0.25
Huber	0.36	0.34	0.34	0.32
Q_n	0.36	0.34	0.33	0.31

Conclusion:

In this study, we used robust heritability estimation methodologies in this study to estimate SNP-based heritability when observations in the dataset were contaminated with outlying observations. For estimating variance components using LMMs, the robust DF-REML estimate method has been described. For the estimation technique and the derivation of the coefficients of determination, Huber's (1964) and Rousseeuw and Croux's (1993) robust estimations of scale were used. The robust approaches were applied to a real dataset, and several procedures were followed, such as estimating heritability and variance components using both original and contaminated data (1%, 3% and 5%). Finally the robust approach is compared with classical approach. It is seen that for all cases heritability estimates decrease as contamination percentage increases.

Development of heritability estimation procedure in case of correlated random effect as well as error

2. Introduction:

In order to study the statistical properties of the genetic parameters one of the basic requirements was to simulate the statistical –biological models with the known population parameters. Basically there are two approaches in the estimation of heritability, one based on the regression of offspring on parents and other based on sib correlation or sib analysis. Both the approaches have their own merits and demits. The precision and bias of these estimators depend on the relationship chosen. In the present we confine only to the estimation of heritability by sib analysis. The generation of data for correlated and uncorrelated cases are undertaken in one-way and two way classification model as discussed below.

2.1 Models for Heritability

2.1.1Introduction

Two important simulation models, namely one way and two way nested models(Half- Sibs and Full – Sibs models) for the estimation of heritability given by Roningen(1974) are used in this study. A brief discussion of these Monte Carlo methods are as follows:

2.1.2 One Way Classification

The one-way classification or half-sib analysis model can be written as follows:

$y_{ij} = \mu + s_i + e_{ij}$; i=1,2,…,s; j=1,2,…,p.

where, y_{ij} is the observed value on the progeny of the j^{th}dam mated to the i^{th}sire,

μ is the general mean,

s_i is the effect due to i^{th}sire,

e_{ij} is the random effect associated with j^{th}progeny of the i^{th}sire.

The simulation model (Ronningen, 1974) to generate half-sib model will be carried out as follows:

$$y_{ij} = \mu + \sigma_s a' + \sigma_e a'_{ij}$$

Where a' and a'_{ij} are random standard normal values.

The value of heritability is given by

$$h^2 = \frac{4\sigma_s^2}{\sigma_s^2 + \sigma_e^2} \qquad \sigma_s^2 = \text{sire variance components}, \quad \sigma_e^2 = \text{error variance components}$$

Correlated Case:

Suppose that sires are independent but within sire progenies are correlated. Further, assume that the correlated errors follow AR(1) i.e.,

$$e_{ij} = \rho e_i(j-1) + \eta_{ij}, \eta_{ij} = \text{random error componets},$$

(1)

$$\text{where} |\rho| < 1, \text{Var}(\eta_{ij}) = \frac{\sigma_e^2}{1-\rho^2} \quad \text{and} \quad \eta_{ij} \approx IIDN(0.1) \quad for\ j > 1.$$

Generate e_{ij} using the equation (1). Then, we can generate the correlated observations $y_{ij}'s$ by using the following modified simulation model:

$$y_{ij} = \mu + \sigma_s a_i' + \sigma_e e_{ij}$$

Notations have the same meaning as defined above, and $e_{ij}'s$ are the values generated from equation (1).

In case of AR(2)

$$e_{ij} = \rho_1 e_i(j-1) + \rho_2 e_i(j-2) + \eta_{ij}, \eta_{ij} = \text{random error componets}$$

In the similar fashion, we can generate the correlated data for different error structures other than AR(1) e.g., AR(2), function of a distance, etc.

2.1.3. Two way nested Model

This is popularly known as full-sib analysis model and can be written as

$$Y_{ijk} = \mu + s_i + d_{ij} + e_{ijk}$$

where s_i = the effect of ithsire

d_{ij} = the effect of jthdam mated to ith sire

e_{ijk} = the random effect associated with kth member of the ijthfull-sib group.

The simulation model used for generating full sib data is given as follows:

$$Y_{ijk} = \mu + \sigma_s a_i + \sigma_d a_{ij} + \sigma_e a_{ijk}$$

Where a_i, a_{ij} and a_{ijk} are standard normal variates.

Results and discussions:

Data have generated using different heritability values i.e. high and low(0.5, 0.1) Half sib AR(!):and different sample size 100 and 500 and different correlation of errors A(1) and AR(2). ρ=-1 to +1. After generating the data variance components are estimated using SAS varcomp proc. ANOVA,ML,REML and MIVQUE methods are used. Estimates of heritability are obtained using the formulae give in the last chapter. Results are tabulated in different tables.

5.1 Half sib Estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for different parametric value of heritability

The data generated from population with low and high heritability for various sample sizes and family structures. The heritability estimates along with MSE(Means Square Error) are obtained and shown in Table1 and Table2. It is seen that value of estimates of heritability changes negative to positive when ρ changes from -1 to +1. MSE value decreasing up to ρ=0, then again increasing ρ is positive. For ρ= -1 to ρ= -0.5 the estimate of heritability is 0 in case of ML, REML and MIVQUE methods and MSE values are not changing. Estimate value of heritability are increasing from ρ=-.4 to ρ=1 .MSE values are showing the same trend. Increasing sample sizes it is noticed that the MSE values are decreasing.

2.2Halfsib Estimate of heritability and MSE values in case of correlated errors (AR(2)) and different sample sizes for different parametric value of heritability

The data generated from population with low and high heritability for various sample sizes and family structures. The heritability estimates along with MSE(Mean Square Error) are obtained and shown in Table3and Table4. It is noticed that in case of AR(2) if fixing AR(1) value changing AR(2) values, the MSE value decrease with increasing the correlation value in general. Sometimes haphazard trends are noticed. We found some good combination of AR(1) and AR(2) values (-0.6,,0.4),(-0.5,-0.1) , (0, 0) ,(0.1,0.1) and (0.1, 0.5) combinations are giving better estimates of heritability. Increasing sample sizes it is noticed that the MSE values are decreasing.

Table 1: Half sib Estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for parametric value of heritability 0.10

MSE								
Rho	ANOVA		ML		REML		MIVQUE	
	P=100	P=500	P=100	P= 500	P=100	P=500	P=100	P=500
-1	0.2293	0.2311	0.0095	0.0095	0.0095	0.0095	0.0095	0.0095
(ht)	-0.3811	-0.3830	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.9	0.2041	0.2067	0.0095	0.0095	0.0095	0.0095	0.0095	0.0095
(ht)	-0.3539	-0.3566	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.8	0.1731	0.1764	0.0095	0.0095	0.0095	0.0095	0.0095	0.0095
(ht)	-0.3177	-0.3215	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.7	0.1416	0.1454	0.0095	0.0095	0.0095	0.0095	0.0095	0.0095
(ht)	-0.2773	-0.2819	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.6	0.112	0.116	0.010	0.010	0.010	0.010	0.010	0.010
(ht)	-0.2343	-0.2399	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-0.5	0.0845	0.0887	0.0095	0.0095	0.0095	0.0095	0.0095	0.0095
(ht)	-0.189	-0.1954	0	0	0	0	0	0
-0.4	0.0601	0.0642	0.0095	0.0095	0.0094	0.0094	0.0094	0.0094
(ht)	-0.1408	-0.1481	0.0004	0.0004	0.0005	0.0005	0.0005	0.0005
-0.3	0.0392	0.0642	0.0092	0.0095	0.0091	0.0094	0.0091	0.0094
(ht)	-0.0889	-0.1481	0.0024	0.0004	0.0032	0.0005	0.0032	0.0005
-0.2	0.0226	0.0258	0.0078	0.0080	0.0076	0.0079	0.0076	0.0079
(ht)	-0.0320	-0.0408	0.0163	0.0149	0.0186	0.0171	0.0186	0.0171
-0.1	0.0118	0.0142	0.0065	0.0068	0.0064	0.0068	0.0064	0.0068
(ht)	0.0314	0.0220	0.0465	0.0430	0.0513	0.0476	0.0513	0.0476
0	0.0095	0.0109	0.0078	0.0078	0.0083	0.0083	0.0083	0.0083
(ht)	0.1037	0.0939	0.1006	0.0958	0.1082	0.1031	0.1082	0.1031
0.1	0.0201	0.0204	0.0180	0.0178	0.0200	0.0197	0.0200	0.0197
(ht)	0.1879	0.1779	0.1792	0.1715	0.1887	0.1806	0.1887	0.1806
0.2	0.0515	0.0506	0.0472	0.0464	0.0515	0.0506	0.0515	0.0506
(ht)	0.2884	0.2784	0.2775	0.2681	0.2884	0.2787	0.2884	0.2787
0.3	0.1176	0.1157	0.1096	0.1079	0.1176	0.1157	0.1176	0.1157
(ht)	0.4114	0.4018	0.3991	0.3896	0.4114	0.4018	0.4114	0.4018
0.4	0.2434	0.2409	0.2301	0.2279	0.2434	0.2409	0.2434	0.2409
(ht)	0.5659	0.5570	0.5520	0.5432	0.5659	0.5570	0.5659	0.5570
0. 5	0.4753	0.4733	0.4542	0.4524	0.4753	0.4733	0.4753	0.4733
(ht)	0.7647	0.7571	0.7490	0.7415	0.7647	0.7571	0.7647	0.7571
0.6	0.8976	0.8980	0.8651	0.8656	0.8976	0.8980	0.8976	0.8980
(ht)	1.0252	1.0195	1.0077	1.0021	1.0252	1.0195	1.0252	1.0195
0.7	1.6534	1.6585	1.6050	1.6103	1.6534	1.6585	1.6534	1.6585

(ht)	1.3663	1.3632	1.3472	1.3442	1.3663	1.3632	1.3663	1.3632
0.8	2.9268	2.9386	2.8600	2.8719	2.9268	2.9386	2.9268	2.9386
(ht)	1.7950	1.7948	1.7752	1.7750	1.7950	1.7948	1.7950	1.7948
0.9	4.7491	4.7670	4.6678	4.6858	4.7498	4.7670	4.7491	4.7670
(ht)	2.268	2.2703	2.2491	2.2515	2.268	2.2703	2.268	2.2703
1	6.5406	6.5607	6.4551	6.4753	6.5406	6.5607	6.5406	6.5607
(ht)	2.6502	2.6538	2.6333	2.6370	2.6502	2.6538	2.65024	2.6538

Table 2 :Half-sib Estimate of heritability and MSE values in case of correlated errors (AR(2)) and different sample sizes in case heritability of 0.10

MSE									
P_1	P_2	ANOVA		REML		ML		MIVQUE	
		P=100	P=200	P=100	P= 200	P=100	P=200	P=100	P=200
-1	-1	0.2460	0.2428	0.1093	0.1093	0.1095	0.10927	0.1095	0.1093
(ht)		-0.2484	-0.2383	0.0023	0.0022	0.0074	0.00842	0.0074	0.0084
-1	-0.5	0.2123	0.2114	0.1095	0.1098	0.1110	0.11177	0.1110	0.1118
(ht)		-0.1738	-0.1589	0.0014	0.0047	0.0132	0.02025	0.0132	0.0202
-1	0.1	0.2113	0.2108	0.1100	0.1103	0.1121	0.11297	0.1121	0.1130
(ht)		-0.1648	-0.1510	0.0035	0.0075	0.0170	0.02418	0.0170	0.0242
-1	1	1.6856	1.5547	1.3008	1.1860	1.6802	1.55046	1.6802	1.5505
(ht)		1.0906	1.0491	0.9288	0.8041	1.0946	1.06020	1.0463	1.0602
-1	0.5	0.2081	0.2085	0.1115	0.1119	0.1155	0.11657	0.1155	0.1166
(ht)		-0.1359	-0.1261	0.0128	0.0161	0.0325	0.03861	0.0325	0.0386
-1	0	0.2114	0.2108	0.1099	0.1101	0.1118	0.11272	0.1118	0.1127
(ht)		-0.1865	-0.1724	0.0031	0.0069	0.0061	0.02332	0.0161	0.0233
-0.8	-0.8	0.1813	0.1866	0.1128	0.1130	0.1200	0.12220	0.1200	0.1222
(ht)		-0.0561	-0.0626	0.0285	0.0381	0.0594	0.07405	0.0394	0.0740
-0.6	0.4	0.1898	0.2026	0.1266	0.1326	0.1472	0.15760	0.1472	0.1576
(ht)		0.0318	0.0454	0.0752	0.0889	0.1256	0.14283	0.1256	0.1428
-0.5	-0.5	0.1800	0.1920	0.1196	0.1236	0.1359	0.14411	0.1359	0.1441
(ht)		0.0074	0.0239	0.0628	0.0777	0.1284	0.14865	0.1284	0.1286
-0.5	-0.1	0.1818	0.1954	0.1220	0.1269	0.1398	0.14960	0.1398	0.1496
(ht)		0.0187	0.0359	0.0674	0.0838	0.1152	0.13602	0.1152	0.1360
-0.5	0.9	1.1004	1.0563	0.8026	0.7701	1.0954	1.05161	1.0954	1.0516
(ht)		0.8736	0.8557	0.7282	0.7104	0.8864	0.86772	0.8764	0.8677
-0.5	0.1	0.1834	0.1074	0.1233	0.1287	0.1415	0.15201	0.1415	0.1520
(ht)		0.0219	0.0387	0.0685	0.0850	0.1176	0.13779	0.1176	0.1378
-0.5	0	0.1825	0.1063	0.1226	0.1277	0.1405	0.15069	0.1405	0.1507
(ht)		0.0202	0.0372	0.0678	0.0843	0.1163	0.13684	0.1163	0.1368
-0.3	0.2	0.1850	0.2008	0.1249	0.1317	0.1444	0.15685	0.1444	0.1569

(ht)		0.0309	0.0486	0.0726	0.0903	0.1229	0.14424	0.1229	0.1442
0	0.5	0.1968	0.2135	0.1323	0.1417	0.1574	0.17260	0.1574	0.1726
(ht)		0.0653	0.0807	0.0921	0.1075	0.1513	0.16892	0.1513	0.1689
0	-0.5	0.1785	0.1945	0.1196	0.1258	0.1360	0.14798	0.1360	0.1480
(ht)		0.0124	0.0328	0.0638	0.0827	0.1101	0.14345	0.1101	0.1345
0	1	3.4235	3.3116	2.7851	2.6810	3.4232	3.31139	3.4232	3.3114
(ht)		1.7385	1.7171	1.5444	1.5209	1.7396	1.71769	1.7396	1.7177
0	-1	0.2063	0.2085	0.1095	0.1107	0.1113	0.11391	0.1113	0.1139
(ht)		-0.1594	-0.1635	0.0036	0.0096	0.0181	0.02844	0.0181	0.0284
0	0	0.1818	0.1990	0.1232	0.1302	0.1415	0.15478	0.1415	0.1548
(ht)		0.0255	0.0461	0.0691	0.0891	0.1177	0.14282	0.1177	0.1428
0.1	0.1	0.1823	0.2000	0.1237	0.1312	0.1421	0.15604	0.1421	0.1560
(ht)		0.0258	0.0476	0.0688	0.0896	0.1174	0.14368	0.1174	0.1437
0.1	0.5	0.1960	0.2132	0.1319	0.1417	0.1568	0.17244	0.1568	0.1724
(ht)		0.0641	0.0802	0.1114	0.1271	0.1503	0.14865	0.1503	0.1687
0.3	-0.2	0.1791	0.1965	0.1215	0.1290	0.1386	0.15284	0.1386	0.1528
(ht)		0.0187	0.0436091	0.0650	0.0869	0.1118	0.14012	0.1118	0.1401

Table 3: Halfsib Estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes parametric value of heritability 0.5

	MSE							
ρ	ANOVA		ML		REML		MIVQUE	
	P=500	p=100	P=500	p=100	P=500	p=100	P=500	p=100
-1	0.023877	0.0234756	0.024161	0.0249525	0.023128	0.023225	0.023867	0.023982
(ht)	0.445295	-0.288932	0.432653	0	0.445295	0	0.445295	0
-0.8	0.364419	0.3774184	0.247697	0.247914	0.247569	0.2477293	0.247569	0.247729
(ht)	-0.100169	-0.110188	0.001928	0.002351	0.00239	0.0028794	0.00239	0.002879
-0.6	0.182937	0.1956736	0.1764717	0.1854192	0.171117	0.1807118	0.171117	0.1807118
(ht)	0.083117	0.0695232	0.086983	0.0771334	0.0940338	0.0833611	0.0740338	0.0833611
-0.4	0.0814276	0.0917786	0.0859963	0.0960323	0.0810401	0.09088	0.0810401	0.09088
(ht)	0.2404321	0.2240275	0.2306491	0.2148756	0.2408062	0.2248447	0.2408062	0.2248447
-0.2	0.0332794	0.0401821	0.0358849	0.0428798	0.0332794	0.0399805	0.0332794	0.0399805
(ht)	0.3789856	0.3606091	0.3670671	0.3491262	0.3789856	0.3608014	0.3789856	0.3608014
0	0.0233299	0.0262235	0.022699	0.0260297	0.0233299	0.0262235	0.0233299	0.0262235
(ht)	0.512632	0.4934539	0.4992923	0.4934539	0.512632	0.4934539	0.512632	0.4934539
0.2	0.0455013	0.0540885	0.1421093	0.0494531	0.0530136	0.0540885	0.0520136	0.0540885
(ht)	0.8933214	0.6456959	0.8765713	0.6311033	0.8933214	0.6456959	0.8933214	0.6456959
0.4	0.0408098	0.1729434	0.0453822	0.1607734	0.0408098	0.1729434	0.0408098	0.1729434
(ht)	0.2983176	0.8612289	0.2872631	0.8448825	0.2983176	0.8612289	0.2983176	0.8612289
0.6	0.1033029	0.5898351	0.1095355	0.5626328	0.1033029	0.5898351	0.1033029	0.5898351
(ht)	0.1787041	1.2307069	0.1691348	1.2122075	0.1787041	1.2307069	0.1787041	1.2307069
0.8	0.1388044	2.011267	1.9755056	1.9567327	0.1388044	2.011267	0.1388044	2.011267
(ht)	0.1274907	1.8967213	0.1185919	1.8770366	0.1274907	1.8967213	0.1274907	1.8967213
1	4.7689802	4.7782916	4.6970401	4.7064034	4.7689802	4.7782916	0.1689802	0.249825
(ht)	2.6779269	2.6796665	2.6612699	2.6630292	2.6779269	2.6796665	2.6779269	0

Table 4: Halfsib Estimate of heritability and MSE values in case of correlated errors (AR(2)) and different sample sizes in case heritability of 0.50

		MSE							
P_1	P_2	ANOVA		ML		REML		MIVQUE	
		P=100	P=200	P=100	P= 200	P=100	P=200	P=100	P=200
-1 (ht)	-1	0.488917 -0.061335	0.5055455 -0.071945	0.3217234 0.033848	0.3192378 0.0346013	0.3085184 0.0625768	0.3033904 0.0658638	0.3085184 0.0625768	0.3033904 0.0658638
-1 (ht)	-0.5	0.3874837 0.0541088	0.4094116 0.041723	0.2940606 0.0912773	0.2924512 0.0941407	0.2804843 0.1413845	0.2794069 0.1446804	0.2804843 0.1413845	0.2794069 0.1446804
-1 (ht)	0.1	0.3789359 0.0694036	0.4013475 0.054783	0.2902304 0.0995042	0.2883118 0.1014682	0.2753359 0.1539653	0.2743213 0.1554791	0.2753359 0.1539653	0.2743213 0.1554791
-1 (ht)	1	1.2395441 1.212535	1.1178785 1.1554846	0.9417583 1.0454333	0.8370438 0.9927172	1.23188 1.2188096	1.105354 1.1653236	1.23188 1.2188096	1.105354 1.1653236
-1 (ht)	0.5	0.3698658 0.3977362	0.392958 0.0783994	0.2861523 0.1163135	0.2843872 0.115657	0.2713197 0.1776563	0.2705514 0.1752478	0.1613197 0.1676563	0.2705514 0.1752478
-1 (ht)	0	0.3799107 0.0671508	0.4021615 0.0530114	0.2905735 0.0983257	0.2888736 0.1003191	0.276074 0.1518926	0.2750476 0.1537561	0.276074 0.1518926	0.2750476 0.1537561
-0.8 (ht)	-0.8	0.2903288 0.2614686	0.3182769 0.232986	0.2531953 0.2078014	0.2543523 0.2033567	0.2442571 0.2989456	0.2497993 0.2881403	0.2442571 0.2989456	0.2497993 0.2881403
-0.6 (ht)	0.4	0.2826397 0.4343475	0.3152194 0.3975651	0.2403634 0.3355253	0.2500989 0.3229166	0.2561565 0.4556018	0.2715525 0.4331193	0.2561565 0.4556018	0.2715525 0.4331193
-0.5 (ht)	-0.5	0.2787143 0.4059834	0.3077416 0.3726499	0.236489 0.3159194	0.2466598 0.3027095	0.2492345 0.4304707	0.2632248 0.4088362	0.2492345 0.4304707	0.2632248 0.4088362
-0.5 (ht)	-0.1	0.2792245 0.4276538	0.3102599 0.3940909	0.2352523 0.3327088	0.2468307 0.3196436	0.2523936 0.449757	0.2682417 0.4282556	0.2523936 0.449757	0.2682417 0.4282556
-0.5 (ht)	0.9	0.9536738 1.1322028	0.8881722 1.0788252	0.6875301 0.9630872	0.6432396 0.9135354	0.9428112 1.1409783	0.8767566 1.0880374	0.9428112 1.1409783	0.8767566 1.0880374
-0.5 (ht)	0.1	0.2791354 0.432518	0.3116034 0.3984701	0.2355222 0.3359199	0.2472748 0.3234566	0.2536769 0.4531939	0.2703652 0.4319457	0.2536769 0.4531939	0.2703652 0.4319457
-0.5 (ht)	0	0.2790786 0.4301217	0.3108332 0.396394	0.2352239 0.3344087	0.2469746 0.3216225	0.2529008 0.4515585	0.2692483 0.4301811	0.2529008 0.4515585	0.2692483 0.4301811
-0.3 (ht)	0.2	0.2814987 0.4480096	0.3158521 0.4144742	0.2367753 0.3474886	0.2495999 0.3360006	0.257519 0.467428	0.2761758 0.4467596	0.257519 0.467428	0.2761758 0.4467596
0 (ht)	0.5	0.2956891 0.4782523	0.3281732 0.4428285	0.2442001 0.373028	0.2558624 0.3595952	0.2717972 0.4980278	0.2879008 0.4759263	0.2717972 0.4980278	0.2879008 0.4759263
0 (ht)	-0.5	0.2790868 0.419876	0.3117267 0.3901806	0.2364726 0.3261852	0.2491653 0.3161415	0.2518388 0.4428596	0.2698825 0.4243884	0.2518388 0.4428596	0.2698825 0.4243884
0 (ht)	1	2.6243948 1.8872541	2.4941297 1.8343796	2.0959868 1.6908519	1.9766438 1.6423128	2.6242349 1.8874116	2.4891787 1.8381763	2.6242349 1.8874116	2.4891787 1.8381763
0 (ht)	-1	0.3512803 0.0844258	0.3785129 0.0687792	0.2821327 0.0983011	0.2816984 0.1026436	0.2601747 0.1581338	0.2625201 0.1615042	0.2601747 0.1581338	0.2625201 0.1615042
0 (ht)	0	0.282824 0.444284	0.3170387 0.4137827	0.2377139 0.3452012	0.2508368 0.3353194	0.2588365 0.4642377	0.2777029 0.4459368	0.2588365 0.4642377	0.2777029 0.4459368
0.1 (ht)	0.1	0.2840212 0.4442807	0.3185621 0.4145428	0.2395047 0.3447372	0.2520232 0.3360139	0.2611644 0.4632117	0.2794824 0.4464846	0.2611644 0.4632117	0.2794824 0.4464846
0.1 (ht)	0.5	0.296347 0.4761647	0.3285226 0.4417194	0.2453556 0.3711972	0.2563454 0.3586618	0.272962 0.4955892	0.2884083 0.4746562	0.272962 0.4955892	0.2884083 0.4746562
0.3 (ht)	-0.2	0.2849197 0.4296791	0.3177481 0.4046221	0.2417599 0.3335078	0.2528274 0.3279882	0.2619686 0.4491103	0.2788514 0.4365253	0.2619686 0.4491103	0.2788514 0.4365253

0.5 (ht)	0.5	0.3036513 0.4685222	0.3305955 0.4459214	0.2566313 0.3627324	0.2623538 0.3589103	0.2858575 0.4833918	0.295549 0.4746578	0.2858575 0.4833918	0.295549 0.4746578
0.5 (ht)	-0.5	0.2885923 0.4091254	0.3172444 0.3932704	0.2484051 0.1264	0.2570623 0.3165383	0.266052 0.4283446	0.2795664 0.4243574	0.266052 0.4283446	0.2795664 0.4243574
0.5 (ht)	0	0.2936112 0.4327986	0.3220002 0.416346	0.2528368 0.3332121	0.2595727 0.3342653	0.2750082 0.4484656	0.2871999 0.4450301	0.2750082 0.4484656	0.2871999 0.4450301
0.6 (ht)	-0.4	0.2998758 0.4299095	0.3230446 0.422885	0.2568548 0.3326517	0.2630846 0.3377191	0.2812265 0.4460873	0.2922839 0.4485868	0.2812265 0.4460873	0.2922839 0.4485868
1 (ht)	0	2.2652188 1.7376081	2.5328067 1.8491667	1.777259 1.547399	2.015278 1.6560534	2.2597589 1.7419088	2.5291173 1.872141	2.2597589 1.7419088	2.5291173 1.852141
1 (ht)	0.5	2.2945728 1.7477899	4.4698334 2.4391805	1.8049469 1.5570371	3.8228249 2.258803	2.2913326 1.7504548	4.4692683 2.4396932	2.2913326 1.7504548	4.4692683 2.4396932
1 (ht)	1	4.4170039 2.4444643	1.2540997 1.2343611	3.7461251 2.2615587	0.946702 1.0669236	4.4158735 2.4454899	1.2436536 1.2327839	4.4158735 2.4454899	1.2436536 1.2427839
1 (ht)	-1	1.103853 1.133445	2.4429774 1.8138037	0.8195423 0.9771187	1.9379834 1.6210762	1.0883625 1.1557986	2.4391544 1.8168861	1.0883625 1.1457986	2.4391544 1.8168861
1 (ht)	-0.5	2.1795182 1.7037757	0.5055455 -0.071945	1.7034232 1.5141305	0.3192378 0.0346013	2.1736527 1.7183973	0.3033904 0.0658638	2.1736527 1.7083973	0.3033904 0.0658638
-1 (ht)	-1	0.488917 -0.061335	0.4094116 0.041723	0.3217234 0.033848	0.3924512 0.0941407	0.3085184 0.0625768	0.2794069 0.1446804	0.3085184 0.0625768	0.2794069 0.1446804

Table 5 : Full sib Estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for parametric value of heritability 0.10

		MSE ;							
Rho	Method	ANOVA		ML		REML		MIVQUE	
	Sample	P=500	P=100	P=500	P= 100	P=500	P=100	P=500	P=100
-1 (ht)	h_s^2	0.098271 -0.181885	0.0983139 -0.183327	0.0152324 0.0014007	0.0153534 0.0002801	0.0152362 0.0014926	0.0153524 0.0003057	0.0152362 0.0014926	0.0153524 0.0003057
	h_d^2	0.9578841 -0.383885	0.9598772 -0.383327	0.6379551 0.0015007	0.639591 0.0002901	0.6378347 0.0914926	0.6395554 0.0013057	0.6378347 0.001526	0.6395554 0.0023057
	h_{s+d}^2	1.1207918 -0.28 8885	1.1210013 -0.283327	0.76337 0.0015307	0.765174 0.0003001	0.7632358 0.0024926	0.7651345 0.0004057	0.7632358 0.0014896	0.7651345 0.0013057
-0.7 (ht)	h_s^2	0.0379526 0.2647565	0.0389461 0.255177	0.0305571 0.2545453	0.0314585 0.2460257	0.0378693 0.2650561	0.0345774 0.2563335	0.0338693 0.2650561	0.0345774 0.2563335
	h_d^2	0.3046345 0.3947565	0.3145789 0.295177	0.3150439 0.2545453	0.3234633 0.2460257	0.3041462 0.2775056	0.3126467 0.2563335	0.3111462 0.3650561	0.3126467 0.2563335
	h_{s+d}^2	0.390546 0.28257565	0.4019273 0..2815517	0.4024871 0.2545453	0.4121844 0.2460257	0.3900127 0.2850567	0.3998217 0.2563335	0.3900127 0.2695561	0.3998217 0.2563335
-0.5 (ht)	h_s^2	0.1491833 0.5039313	0.1595625 0.4903359	0.1488045 0.4906877	0.1495852 0.4773262	0.1691833 0.5039313	0.1595402 0.4904137	0.1591833 0.5039313	0.1595402 0.4904137
	h_d^2	0.1125158 0.6037113	0.121275 0.5903359	0.1200414 0.5906877	0.128886 0.4773262	0.1125158 0.5139314	0.1211475 0.4904137	0.1125158 0.5989313	0.1211475 0.5904137
	h_{s+d}^2	0.1625511 0.5145313	0.1733496 0.5090335	0.1720633 0.5986877	0.1829121 0.4773262	0.1625511 0.5239317	0.1732105 0.4904137	0.1625511 0.5187313	0.1732105 0.5104137
-0.3 (ht)	h_s^2	0.3455414 0.6860825	0.9612935 1.0813578	0.3284329 0.6711132	0.9270483 1.0635749	0.3455414 0.6860825	0.9612935 1.0813578	0.3455414 0.6860825	0.3612935 1.0813578
	h_d^2	0.0426168 0.8860825	0.1239811 1.5813578	0.0457459 0.6811132	0.1137773 2.0635882	0.0426168 0.9863581	0.1239811 2.0813578	0.0426168 0.6860825	0.1239811 2.0814579
	h_{s+d}^2	0.0653294 0.8860825	0.0874024 1.0983578	0.0707039 0.6921132	0.0798661 1.5635231	0.0653294 0.7860831	0.0874024 1.6781359	0.0653294 0.6860825	0.0874024 1.3893576
0 (ht)	h_s^2	0.6271941 0.893568	0.603396 0.8764351	0.6014082 0.8769554	0.5783213 0.8599561	0.6271941 0.893568	0.603396 0.8764351	0.6271941 0.893478	0.603396 0.8764351
	h_d^2	0.043762 0.983568	0.0431264 1.8014351	0.0404352 0.8769554	0.0403303 1.8599332	0.043762 1.993535	0.0431264 1.8764345	0.043062 1.803216	0.0431264 1.8764321
	h_{s+d}^2	0.0353518 0.895568	0.0372861 0.9884351	0.0345169 0.8769554	0.0369619 0.9599542	0.0353518 1.043569	0.0372861 0.9964323	0.035351 0.993568	0.0372861 0.9764363
0.3 (ht)	h_s^2	0.9853297 1.096046	0.9578041 1.0798574	0.9504265 1.0781632	0.9236276 1.0620805	0.9853297 1.096046	0.9578041 1.0798574	0.9453297 1.096046	0.9578041 1.0798574
	h_d^2	0.1281597	0.1225202	0.117433	0.1123769	0.1281597	0.1225202	0.1201597	0.1225202

		2.096046	1.1798574	1.0791632	2.0620836	1.106046	1.0798574	1.066046	1.0798574
	h_{s+d}^2	0.0893778 1.196046	0.0861666 1.1985741	0.0813335 1.07781632	0.0786898 1.5627685	0.0893778 1.119604	0.0861666 1.0798574	0.0853778 1.096046	0.0861666 1.0798574
0. 5 (ht)	h_s^2	1.4527074 1.310457	1.4237997 1.2956129	1.4079084 1.2915827	1.379705 1.2768211	1.4527074 1.310457	1.4237997 1.2956129	1.4227074 1.310457	1.4237997 1.2956129
	h_d^2	0.3056672 1.810457	0.2968278 1.3956129	0.286385 1.2934827	0.2781384 2.2768342	0.3056672 2.315468	0.2968278 1.2956129	0.3056672 2.317894	0.3168278 1.2956129
	h_{s+d}^2	0.2347236 1.420457	0.2281108 1.2985129	0.2182726 1.2978827	0.2122402 1.8768215	0.2347236 1.587567	0.2281108 1.2956129	0.2347236 1.657892	0.2381108 1.2956129
0.7 (ht)	h_s^2	2.5966888 1.7202475	2.5703997 1.7088652	2.5339041 1.7004875	2.5082108 1.6891521	2.5966888 1.7202475	2.5703997 1.7088652	2.5466888 1.7202475	2.5703997 1.7088652
	h_d^2	0.8956373 2.1202475	0.8847362 1.7288652	0.8595667 1.7004875	0.8491983 2.6891534	0.8956373 2.7202475	0.8847362 1.7088652	0.8756373 2.7292471	0.8847362 2.7088652
	h_{s+d}^2	0.7632251 1.8902475	0.7540315 1.7788652	0.7301186 1.7004875	0.7214505 1.9891567	0.7632251 1.8802462	0.7540315 1.7088652	0.7432251 1.8892476	0.7540315 1.7988865
1 (ht)	h_s^2	6.7818181 2.7242263	6.8135921 2.78902	6.6974755 2.7078968	6.7295237 2.7127551	6.7818181 2.7242263	6.8135921 2.72902	6.7818181 2.7242263	6.8135921 2.72902
	h_d^2	3.7234494 2.9842263	3.7487426 2.9810223	3.6611832 2.8178968	3.6866634 3.7127583	3.7234494 3.7242973	3.7487426 3.756935	3.7234494 3.7242289	3.7487426 3.72673

Table 6: Full sib Estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for parametric value of heritability 0.50

MSE									
Rho	Method	ANOVA		ML		REML		MIVQUE	
	Sample	500	100	500	100	500	100	500	100
-1 (ht)	h_s^2	0.4745049 0.072369	0.4821838 0.078842	0.3576657 0.0136716	0.3628984 0.0084092	0.3563739 0.0149473	0.3615089 0.0096753	0.3563739 0.0149473	0.3615089 0.0096753
	h_d^2	0.7696736 0.082369	0.7798103 0.088842	0.6201683 0.0336716	0.6273989 0.0184092	0.6183922 0.0299473	0.6255288 0.0196754	0.6183987 0.2149465	0.6255288 0.0296765
	h_{s+d}^2	1.684612 0.080469	1.7002153 0.098842	1.4624506 0.0134716	1.4741249 0.0086092	1.4595972 0.0178647	1.4711857 0.0099975	1.4595972 0.0199478	1.4711857 0.0099798
-0.7 (ht)	h_s^2	0.0362448 0.5215111	0.0402199 0.501621	0.038248 0.5081017	0.0426125 0.4884681	0.0362448 0.5215111	0.0401717 0.5016599	0.0362448 0.5215111	0.0401717 0.5016599
	h_d^2	0.1059406 0.5675111	0.1174672 0.541621	0.1130348 0.5381017	0.1248533 0.4984681	0.1059406 0.9265113	0.1174041 0.9501659	0.1059406 1.5215784	0.1174041 1.50187534
	h_{s+d}^2	0.5193806 0.5215111	0.5477033 0.531621	0.5377983 0.5181017	0.5661963 0.4804681	0.5193806 0.6289117	0.5476073 0.5516598	0.5193806 0.5900151	0.5476073 0.6016556
-0.5 (ht)	h_s^2	0.0754066 0.8097071	0.0878031 0.7851935	0.068745 0.7937283	0.0619276 0.7694161	0.0754066 0.8097071	0.0678031 0.7851935	0.0754066 0.8097071	0.0678031 0.7851935
	h_d^2	0.0356855 0.8397071	0.0373889 0.7951935	0.0350905 0.8937283	0.0375035 0.794161	0.0356855 0.8697071	0.0373889 1.1125193	0.0356855 1.6797895	0.0373889 1.4851989
	h_{s+d}^2	0.20576 0.8497071	0.2281638 0.7871935	0.2186582 0.8537283	0.2416015 0.8694161	0.20576 0.8597071	0.2281638 0.7981934	0.20576 0.895432	0.2281638 0.8964321
-0.3 (ht)	h_s^2	0.2017277 1.0127343	0.1839696 0.986424	0.1875438 0.9953401	0.170756 0.9692075	0.2017277 1.0127343	0.1839696 0.986424	0.2017277 1.0127343	0.1839696 0.986424
	h_d^2	0.0849251 1.043343	0.077156 1.786424	0.0773452 0.9953401	0.0704789 0.9692075	0.0849251 2.0127363	0.077156 1.999427	0.0849251 2.0127376	0.077156 1.186421
	h_{s+d}^2	0.0835544 1.0537343	0.0980029 0.9996424	0.0906629 0.9953401	0.1058641 0.9692075	0.0835544 1.3127347	0.0980029 0.9964398	0.0835544 1.0887344	0.0980029 0.999421
0 (ht)	h_s^2	0.4109657 1.2165893	0.3847571 1.191063	0.3884926 1.1981041	0.3633111 1.1727144	0.4109657 1.2165893	0.3847571 1.191063	0.4109657 1.2165893	0.3847571 1.191063
	h_d^2	0.2167674 1.3165812	0.2002501 2.191063	0.2013123 1.1981041	0.1857704 1.1727144	0.2167674 2.2165887	0.2002501 2.191026	0.2167674 2.2165894	0.2002501 2.191066
	h_{s+d}^2	0.0432525 1.4165833	0.0482907 1.291063	0.0434071 1.1981041	0.0493054 1.1727144	0.0432525 1.3465875	0.0482907 1.401069	0.0432525 1.3965896	0.0482907 1.291061
0.3 (ht)	h_s^2	0.6359402 1.3780468	0.6090954 1.3566075	0.6066396 1.3589356	0.5806974 1.3375917	0.6359402 1.3780468	0.6090954 1.3566075	0.6359402 1.3780468	0.6090954 1.3566075
	h_d^2	0.3804428 1.4280464	0.3617376 2.3566075	0.358398 1.3589356	0.3405591 1.3375917	0.3804428 2.3780445	0.3617376 2.3566075	0.3804428 2.3786732	0.3617376 2.4560985
	h_{s+d}^2	0.070586 1.4480321	0.0699851 1.4566075	0.0646795 1.3589356	0.0648645 1.3375917	0.070586 1.4380473	0.0699851 1.4076607	0.070586 1.978046	0.0699851 1.4568347

0. 5 (ht)	h_s^2	0.910936 1.5387524	0.8871555 1.5216589	0.8748306 1.5192238	0.8518017 1.5021967	0.910936 1.5387524	0.8871555 1.5216589	0.910936 1.5387524	0.8871555 1.5216589
	h_d^2	0.5944249 1.6387524	0.5771342 2.5216545	0.5657337 2.5192238	0.5491693 2.5021967	0.5944249 2.5387524	0.5771342 2.5213583	0.5944249 2.5387556	0.5771342 2.5216554
	h_{s+d}^2	0.1488611 1.5688752	0.1460049 1.5983481	0.1366608 1.6192238	0.1344748 1.6421967	0.1488611 1.6587524	0.1460049 1.8212567	0.1488611 1.7687528	0.1460049 1.7216521
0.7 (ht)	h_s^2	1.6325969 1.8683506	1.6172951 1.8580999	1.5833143 1.848602	1.5685335 1.8383872	1.6325969 1.8683506	1.6172951 1.8580999	1.6325969 1.8683506	1.6172951 1.8580999
	h_d^2	1.1909502 1.8883506	1.1795402 2.8580912	1.1491653 2.848602	1.1382628 2.8383872	1.1909502 2.8683513	1.1795402 2.8580999	1.1909502 2.8683507	1.1795402 2.8580991
	h_{s+d}^2	0.4670591 1.8553506	0.4643053 1.8980991	0.4419508 1.899863	0.4396741 1.8999872	0.4670591 1.8989508	0.4643053 1.8980999	0.4670591 1.8983502	0.4643053 1.8980993
1 (ht)	h_s^2	4.612573 2.7529176	4.6506737 2.7603397	4.5441372 2.7368118	4.5824519 2.744318	4.612573 2.7529176	4.6506737 2.7603397	4.612573 2.7529176	4.6506737 2.7603397
	h_d^2	3.8350907 2.9529176	3.8703735 3.7603366	3.7727696 3.7368116	3.8082345 2.744318	3.8350907 3.7529112	3.8703735 3.7603345	3.8350907 3.7529171	3.8703735 3.7603332
	h_{s+d}^2	2.3642319 2.8529176	2.3932471 2.7897323	2.3155112 2.7868112	2.3446376 2.744318	2.3642319 2.9529178	2.3932471 2.9803332	2.3642319 2.9529173	2.3932471 2.8903393

Table 7: Full sib Estimate of heritability and MSE values in case of correlated errors (AR(2)) and different sample sizes in case heritability of 0.10

	METHOD	MSE							
		ANOVA		ML		REML		MIVQUE	
	Sample	100	500	100	500	100	500	100	500
Rc1=-1 Rc2=-1	MSE1 (ht1)	0.159 -0.247	0.157 -0.231	0.015 0.005	0.014 0.009	0.015 0.010	0.013 0.017	0.018 0.017	0.017 0.015
	MSE2 (ht2)	17.018 4.247	16.899 4.231	14.989 3.996	14.956 3.991	14.945 3.989	14.895 3.983	14.895 3.983	14.895 3.983
	MSE3 (ht3)	3.519 2.5	3.324 2.4	3.654 2.6	3.519 2.7	3.543 2.8	3.512 2.4	3.588 2.6	3.467 2.2
Rc1=-1 Rc2=-.5	MSE1 (ht1)	0.122 -0.168	0.120 -0.151	0.015 0.014	0.0169 0.025	0.016 0.029	0.021 0.044	0.021 0.044	0.021 0.044
	MSE2 (ht2)	16.392 4.168	16.273 4.151	14.915 3.986	14.835 3.975	14.809 3.972	14.698 3.956	14.698 3.956	14.697 3.956
	MSE3 (ht3)	3.519 2.7	3.123 2.6	3.413 2.5	3.351 2.6	3.567 2.7	3.534 2.2	3.619 2.5	3.519 2.3
Rc1=0 Rc2=-.5	MSE1 (ht1)	0.085 0.049	0.085 0.049	0.024 0.087	0.024 0.087	0.042 0.138	0.042 0.138	0.042 0.138	0.042 0.138
	MSE2 (ht2)	14.723 3.951	14.722 3.942	14.379 3.912	14.379 3.913	14.014 3.862	14.012 3.762	14.014 3.866	14.011 3.862
	MSE3 (ht3)	3.519 2.5	3.319 2.3	3.534 2.8	3.519 2.6	3.546 2.5	3.522 2.1	3.523 2.8	3.509 2.6
Rc1=0 Rc2=.5	MSE1 (ht1)	0.104 0.103	0.103 0.103	0.038 0.117	0.037 0.114	0.064 0.181	0.063 0.180	0.064 0.181	0.062 0.179
	MSE2 (ht2)	14.339 3.897	14.335 3.892	14.169 3.883	14.166 3.881	13.717 3.819	13.713 3.812	13.715 3.819	13.712 3.813
	MSE3 (ht3)	3.519 2.4	3.565 2.2	3.654 2,3	3.542 2.1	3.519 2.4	3.322 2.2	3.632 2.8	3.519 2.6
Rc1=1 Rc2=.5	MSE1 (ht1)	2.996 1.674	2.357 1.783	2.387 1.484	2.225 1.340	2.995 1.677	2.357 1.434	3.357 1.784	3.257 1.674

	MSE2 (ht2)	5.444 2.326	4.989 2.217	6.261 2.516	5.803 2.410	5.419 2.323	4.975 2.216	4.975 2.245	4.655 2.216
	MSE3 (ht3)	3.519 2.6	3.521 2.5	3.996 2.4	3.987 2.1	3.874 2.5	3.765 2.3	3.519 2.8	3.432 2.5

* ht1 ,ht2 and ht3 are obtained using formulae considering only sire component, only dam component and both sire and dam component same for MSE1,MSE2 and MSE3

Table 8: Full sib Estimate of heritability and MSE values in case of correlated errors (AR(2)) and different sample sizes in case heritability of 0.5

MSE									
		ANOVA		ML		REML		MIVQUE	
		P=100	p=500	P=100	p=500P=100		p=500P=100		p=500
Rc1=-1 Rc2=-1	MSE (ht1)	0.4721022 0.016912	0.5050107 0.018518	0.3201617 0.0649644	0.3161979 0.0800881	0.3003282 0.1002332	0.2967683 0.1201476	0.3003282 0.1002332	0.2967683 0.1201476
	MSE (ht2)	11.684765 4.0169119	11.726601 4.0185178	11.077648 3.9350356	10.989606 3.9199119	10.861744 3.8997668	10.747473 3.8798524	10.861744 3.8997668	10.747473 3.8798524
	MSE3 (ht3)	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2
Rc1=-1 Rc2=-.5	MSE (ht1)	0.345269 0.135154	0.3922886 0.1243975	0.2758018 0.1379398	0.2771179 0.1522402	0.2508905 0.200361	0.2591862 0.2143679	0.2508905 0.200361	0.2591862 0.2143679
	MSE (ht2)	10.712549 3.864846	10.819367 3.8756025	10.627594 3.8620602	10.54941 3.8477598	10.255663 3.799639	10.18609 3.7856321	10.255663 3.799639	10.18609 3.7856321
	MSE3 (ht3)	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2
Rc1=0 Rc2=-.5	MSE (ht1)	0.2095713 0.0492439	0.290843 0.5069488	0.1715443 0.0870818	0.2234997 0.416119	0.1858822 0.1380682	0.2489659 0.5358269	0.1858822 0.1380682	0.2489659 0.5358269
	MSE (ht2)	8.3296764 3.9507561	8.5911957 3.4930512	8.9155416 3.9129182	9.0288042 3.583881	8.2115175 3.8619318	8.3887754 3.4641731	8.2115175 3.8619318	8.3887754 3.4641731
	MSE3 (ht3)	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2
Rc1=0 Rc2=.5	MSE (ht1)	0.2233608 0.5393713	0.2985235 0.5609866	0.1762173 0.4271468	0.2259265 0.4595421	0.2030454 0.5563644	0.2633412 0.5854255	0.2030454 0.5563644	0.2633412 0.5854255
	MSE (ht2)	8.0183033 3.4606287	8.2984624 3.4390134	8.6522973 3.5728532	8.7898279 3.5404579	7.9190841 3.4436356	8.1274165 3.4145745	7.9190841 3.4436356	8.1274165 3.4145745
	MSE3 (ht3)	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 7 2	1.9316288 2	1.9316288 2
Rc1=1 Rc2=.5	MSE (ht1)	0.2233608 0.5978609	2.2715493 1.9058696	0.1762173 0.4753393	1.787283 1.7112097	0.2030454 0.612054	2.2702529 1.9068249	0.2030454 0.612054	2.2702529 1.9068249
	MSE (ht2)	8.0183033 3.4021391	2.7948503 2.0941304	8.6522973 3.5246607	3.3927611 2.2887903	7.9190841 3.387946	2.7882435 2.0931751	7.9190841 3.387946	2.7882435 2.0931751
	MSE3 (ht3)	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2	1.9316288 2

2.3Full sib Estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for different parametric value of heritability

The data generated from population with low and high heritability for various sample sizes and family structures. The heritability estimates along with MSE(Means Square Error) are obtained and given in Table5and Table 6. In almost all cases biased estimates are obtained. Estimates for considering only sire components are better than considering both sire and dam component and dam component alone.

2.4 Full-sib estimate of heritability and MSE values in case of correlated errors (AR(2)) and different sample sizes for different parametric value of heritability

The data generated from population with low and high heritability for various sample sizes and family structures. The heritability estimates along with MSE(Means Square Error) are obtained and shown in Table7 and Table 8. It is noticed In almost all cases biased estimates are obtained. Estimates for considering only sire components are better than considering sire and dam component and dam component alone. Combination of correlation i.e. (0,-.5) and (0, .5) are giving better results than any other combination. Increasing sample size decrease the MSE values.

The random effect model for the one-way classification is-

$$y_{ij} = \mu + \alpha_i + e_{ij} \quad i=1,2,\ldots,a; \quad j=1,2,\ldots,n.$$

where, y_{ij} is the observed value of the ith class, μ is the general mean, α_i is the effect due toi^{th}sire, e_{ij} is the residual error.

Under the normality assumptions

$$E\left(e_{ij}\right)=0,\ E(\alpha_i)=0,\ E\left(e_{ij}^2\right)=\sigma_e^2,\ E\left(\alpha_i^2{}_{ij}\right)=\alpha_\alpha^2,\ Cov(\alpha_i,\alpha_{i'})\neq 0\ \ \forall \mathrm{i}\neq \mathrm{i}',$$

Correlated Case:

Suppose that sires are independent but within sire progenies are correlated. Further, assume that the correlated errors follow AR(1) i.e.,

$$e_{ij} = \rho e_i(j-1) + \eta_{ij}, \eta_{ij} = \text{random error componets}, \quad (1)$$

$$\text{where} |\rho| < 1, \text{Var}(\eta_{ij}) = \frac{\sigma_e^2}{1-\rho^2} \quad \text{and} \quad \eta_{ij} \approx IIDN(0.1) \quad for\ j > 1.$$

Generate e_{ij} using the equation (1). Then, we can generate the correlated observations $y_{ij}'s$ by using the following modified simulation model:

$$y_{ij} = \mu + \sigma_s a_i' + \sigma_e e_{ij}$$

Notations have the same meaning as defined above, and $e_{ij}'s$ are the values generated from equation (1).

We generate sire component and error component following different combination of distribution i.e. normal, beta, Cauchy and t-distribution with different heritability values and estimate of heritability and rmse values obtained by four different methods i.e. ANOVA, ML,REML and MIVQUE methods with different parametric values of heritability.

The results are tabulated here. From the results it is noticed that when we increase correlation value –ve to zero the rmse values decrease . If we increase from zero to higher i,e. nearer to +1 it is noticed that rmse values increased for all the combinations of distribution.

Table 9: Half sib Estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for parametric value of heritability 0.10 when sire is distributed as Normal but error follows Beta distribution

Heritability(Ht)	**Correlation(RC)**	**Estimated Heritability**	**M.S.E**
0.1	-0.7	-0.779	0.780
		0.000	0.009
		0.000	0.009
		0.000	0.009
0.1	-0.3	-0.323	0.215
		0.004	0.005
		0.006	0.007

		0.006	0.007
0.1	0	0.099	0.049
		0.095	0.018
		0.107	0.020
		0.107	0.020
0.1	0.3	0.578	0.287
		0.554	0.263
		0.579	0.287
		0.579	0.287
0.1	0.7	1.315	1.532
		1.288	1.467
		1.315	1.532
		1.315	1.532

Table 10: Half sib Estimate of heritability and MSE values in case of Correlated errors (AR(1)) and different sample sizes for parametric value of heritability 0.10 when sire is distributed as Normal but error follows Cauchy distribution

Heritability(Ht)	**Correlation(RC)**	**Estimated Heritability**	**M.S.E**
0.3	-0.7	-0.740	1.196
		0.000	0.089
		0.000	0.089
		0.000	0.089
0.3	-0.3	-0.427	0.589
		0.005	0.090
		0.006	0.090
		0.006	0.090
0.3	0	0.015	0.095
		0.035	0.076
		0.044	0.072
		0.044	0.072

0.3	0.3	-0.427	0.589
		0.005	0.090
		0.006	0.090
		0.006	0.090
0.3	0.7	1.390	1.634
		1.365	1.579
		1.390	1.634
		1.390	1.634

Table11 : Half sib Estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for parametric value of heritability 0.10 when sire is distributed as t- distribution but error follows t- distribution

Heritability(Ht)	Correlation(RC)	Estimated Heritability	M.S.E
0.9	-0.7	-0.872	3.148
		0.000	0.810
		0.000	0.810
		0.000	0.810
0.9	-0.3	-0.722	2.643
		0.000	0.810
		0.000	0.810
0.9	0	-0.118	1.072
		0.025	0.769
		0.030	0.761
		0.030	0.761
0.9	0.3	0.477	0.231
		0.455	0.247
		0.478	0.228
		0.478	0.228
0.9	0.7	1.326	0.233
		1.300	0.211
		1.326	0.233

		1.326	0.233

Table 12: Half sib Estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for parametric value of heritability 0.90 when sire is distributed as Normal but error follows Normal distribution

Heritability(Ht)	Correlation(RC)	Estimated Heritability	M.S.E
0.9	-0.7	-0.218	1.301
		0.019	0.779
		0.023	0.774
		0.023	0.774
0.9	-0.3	0.349	0.376
		0.340	0.374
		0.361	0.353
		0.361	0.353
0.9	0	0.734	0.113
		0.710	0.120
		0.735	0.112
		0.735	0.112
0.9	0.3	1.318	0.275
		1.292	0.253
		1.318	0.275
		1.318	0.275
0.9	0.7	2.157	1.663
		2.133	1.604
		2.157	1.663
		2.157	1.663

Table 13: Half sib Estimate of heritability and MSE values in case of correlated errors (AR(1)) and different sample sizes for parametric value of heritability 0.90 when sire is distributed as Cauchy distribution but error follows Cauchy distribution

Heritability(Ht)	Correlation(RC)	Estimated Heritability	M.S.E
0.9	-0.7	0.929	3.212
		1.262	2.251
		1.273	2.266
		1.273	2.266
0.9	-0.3	1.672	3.052
		1.741	2.839
		1.755	2.860
		1.755	2.860
0.9	0	2.022	3.403
		2.008	3.381
		2.023	3.401
		2.023	3.401
0.9	0.3	2.431	3.403
		2.413	3.365
		2.431	3.403
		2.431	3.403
0.9	0.7	3.583	7.273
		3.576	7.235
		3.583	7.273
		3.583	7.273

Objective-III

1. Introduction:

Occurrence of outlier(s) is very common in every field in which data collection is involved. An outlier in a set of data is an observation (or an observation vector) that appears to be inconsistent with the remainder of the observations in that data set. Outlier(s) in genetics and breeding experiments is/ are also likely to appear. The impact of outlier in plant breeding is erroneous results in heritability estimation and genomic prediction. Here, our objective is to develop a suitable robust procedure for heritability estimation. The existing methods are not robust against the presence of outlier and non-normality of error or random effect for estimating heritability. There is a need to modify the existing methods or to develop the robust procedure for estimating heritability under different distributional assumptions of the random effect as well as the error term.

Since in most of the practical situations in plant as well as animal breeding data the assumption of independent errors is not fulfilled and they may be correlated. There is little work (no work) towards the development of methods for estimation of heritability in case of correlated error and correlated random effect for common dam as well as common dam and sire. So there is need to develop methodology for estimation of important genetic parameter heritability by considering the correlated random effects and error structure in the models.

The assumptions of uncorrelated error in Henderson's mixed model do not hold always in practical situation. Therefore, there is a need to modify the above model for correlated error as well as random effect. The researchers estimate genetic parameters by assuming independent errors. The consequence of this is the upward or downward biased estimates of genetic parameters depending upon the nature of correlation among errors. So there is a need to develop breeding value estimation procedure using Henderson's Mixed model under correlated errors.

2. Statistical methodology:

The Usual Representation of the Mixed effect linear model is given below:

Let the linear mixed model is represented as

$$Y = X\beta + Zg + e$$

where,

$Y = [Y_1, Y_2 \ldots\ldots\ldots Y_n]'$ is the vector of Observations

$\beta = [\beta_1, \beta_2 \ldots\ldots\ldots \beta_n]'$ is the vector of Fixed effects

$g = [g_1, g_2 \ldots\ldots\ldots g_n]'$ is the vector of random genetic effects

e = $[e_1, e_2 \ldots\ldots\ldots\ldots e_n]'$is the vector of random residual effects

$\boldsymbol{X_{n\times p}}$ = Incidence Matrix relating element of Y to element of β

$\boldsymbol{Z_{n\times p}}$ = Incidence Matrix relating elements of Y to elements of g.

The distribution of fixed and random effects are as follows

$$\begin{bmatrix} Y \\ g \\ e \end{bmatrix} \sim \text{MVN}\left\{ \begin{bmatrix} \boldsymbol{X}\beta \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} ZGZ' + R & ZG & R \\ GZ' & G & 0 \\ R & 0 & R \end{bmatrix} \right\}$$

Maximizing joint density of y and g

f(y,g) = f(y | g) f(g)

Where f(y,g) ~ MVN ($\boldsymbol{X}\beta$ + Zg, R)

f(g) ~ MVN (0, G)

The generalized BLUP estimate of

$\hat{\beta}$ =[$X'V^{-1}X]^{-1}X'V^{-1}$Y

$\hat{g}$= $GZ'V^{-1}$ (Y-X$\hat{\beta}$)

for independent error where R= Iσ_e^2

V(e) = σ_e^2

V($\hat{g}$) = σ_e^2 $GZ'PZG$ [Where P = V^{-1}- $V^{-1}X(X'V^{-1}X)X'V^{-1}$]

V($\hat{y}$) = ZGZ'+ Iσ_e^2

For correlated error Structure

$$\text{R} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1n}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \cdots & \sigma_{nn}^2 \end{bmatrix}$$

V(e) = R

V($\hat{g}$) = $GZ'V^{-1}ZG' - GZ'V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}ZG'$

V($\hat{y}$) = ZGZ' + R

for correlated Random Effects

$$\text{G} = \begin{bmatrix} \sigma_{g1}^2 & \cdots & \sigma_{gq}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{gn}^2 & \cdots & \sigma_{gnq}^2 \end{bmatrix}$$

V(e) =R

V($\hat{g}$) = RGZ'PZG [Where P = V^{-1}- $V^{-1}X(X'V^{-1}X)X'V^{-1}$]

V($\hat{y}$) = ZGZ' + R

For both correlated error and random effects

$$\text{G} = \begin{bmatrix} \sigma_{g1}^2 & \cdots & \sigma_{gq}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{gn}^2 & \cdots & \sigma_{gnq}^2 \end{bmatrix}$$

$$\text{R} = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1n}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \cdots & \sigma_{nn}^2 \end{bmatrix}$$

$$V(e) = R$$
$$V(\hat{g}) = RGZ'PZG$$
$$V(\hat{y}) = ZGZ' + R$$

3. Data description: The data generated from population with low and high heritability for various sample sizes and family structures. Two dataset has been generated, incorporating correlated and uncorrelated error structures. SAS programme for data generation is given in next section.

4. Results and discussion:

The dataset is analyzed and variance component estimation is done using the Henderson's mixed model. Table 1 shows the variance component estimates done by analyzing linear mixed model approach using Henderson's mixed model equation. Results show the mean heritability of the trait of the interest is 0.39. The table also shows the estimated breeding values for all the animals of the study.

Table 1: Comparison of breeding values in correlated and uncorrelated case

Without Correlated			Correlated		
Heritability	Yield	Breeding Value	Heritability	Yield	Breeding Value
0.39	0.55	0.2145	0.32	0.55	0.176
0.39	0.77	0.3003	0.32	0.77	0.2464
0.39	0.77	0.3003	0.32	0.77	0.2464
0.39	0.97	0.3783	0.32	0.97	0.3104
0.39	0.98	0.3822	0.32	0.98	0.3136
0.39	2.73	1.0647	0.32	2.73	0.8736
0.39	0.56	0.2184	0.32	0.56	0.1792
0.39	2.29	0.8931	0.32	2.29	0.7328
0.39	0.59	0.2301	0.32	0.59	0.1888
0.39	0.3	0.117	0.32	0.3	0.096
0.39	2.11	0.8229	0.32	2.11	0.6752
0.39	2.38	0.9282	0.32	2.38	0.7616
0.39	0.52	0.2028	0.32	0.52	0.1664
0.39	1.87	0.7293	0.32	1.87	0.5984
0.39	0.55	0.2145	0.32	0.55	0.176
0.39	0.3	0.117	0.32	0.3	0.096
0.39	1.37	0.5343	0.32	1.37	0.4384
0.39	3.17	1.2363	0.32	3.17	1.0144
0.39	1.26	0.4914	0.32	1.26	0.4032

0.39	0.25	0.0975	0.32	0.25	0.08
0.39	0.68	0.2652	0.32	0.68	0.2176
0.39	0.5	0.195	0.32	0.5	0.16
0.39	0.02	0.0078	0.32	0.02	0.0064
0.39	0.76	0.2964	0.32	0.76	0.2432
0.39	0.98	0.3822	0.32	0.98	0.3136
0.39	0.75	0.2925	0.32	0.75	0.24
0.39	0.79	0.3081	0.32	0.79	0.2528
0.39	0.23	0.0897	0.32	0.23	0.0736
0.39	3.37	1.3143	0.32	3.37	1.0784
0.39	1.44	0.5616	0.32	1.44	0.4608
0.39	0.33	0.1287	0.32	0.33	0.1056
0.39	0.14	0.0546	0.32	0.14	0.0448
0.39	0.46	0.1794	0.32	0.46	0.1472
0.39	0.88	0.3432	0.32	0.88	0.2816
0.39	1.76	0.6864	0.32	1.76	0.5632
0.39	0.56	0.2184	0.32	0.56	0.1792
0.39	0.17	0.0663	0.32	0.17	0.0544
0.39	0.38	0.1482	0.32	0.38	0.1216
0.39	0.01	0.0039	0.32	0.01	0.0032
0.39	0.44	0.1716	0.32	0.44	0.1408
0.39	2.15	0.8385	0.32	2.15	0.688
0.39	0.68	0.2652	0.32	0.68	0.2176
0.39	1.19	0.4641	0.32	1.19	0.3808
0.39	2.11	0.8229	0.32	2.11	0.6752
0.39	0.97	0.3783	0.32	0.97	0.3104
0.39	1.55	0.6045	0.32	1.55	0.496
0.39	0.02	0.0078	0.32	0.02	0.0064
0.39	1.06	0.4134	0.32	1.06	0.3392
0.39	0.69	0.2691	0.32	0.69	0.2208
0.39	0.29	0.1131	0.32	0.29	0.0928
0.39	0.94	0.3666	0.32	0.94	0.3008
0.39	3.08	1.2012	0.32	3.08	0.9856
0.39	0.06	0.0234	0.32	0.06	0.0192
0.39	0.45	0.1755	0.32	0.45	0.144
0.39	1.25	0.4875	0.32	1.25	0.4
0.39	1.56	0.6084	0.32	1.56	0.4992
0.39	1.74	0.6786	0.32	1.74	0.5568
0.39	0.05	0.0195	0.32	0.05	0.016
0.39	0.65	0.2535	0.32	0.65	0.208
0.39	1.15	0.4485	0.32	1.15	0.368

0.39	0.35	0.1365	0.32	0.35	0.112
0.39	0.23	0.0897	0.32	0.23	0.0736
0.39	0.76	0.2964	0.32	0.76	0.2432
0.39	0.67	0.2613	0.32	0.67	0.2144
0.39	0.17	0.0663	0.32	0.17	0.0544
0.39	0.26	0.1014	0.32	0.26	0.0832
0.39	0.67	0.2613	0.32	0.67	0.2144
0.39	0.16	0.0624	0.32	0.16	0.0512
0.39	0.71	0.2769	0.32	0.71	0.2272
0.39	3.83	1.4937	0.32	3.83	1.2256
0.39	0.87	0.3393	0.32	0.87	0.2784
0.39	0.6	0.234	0.32	0.6	0.192
0.39	0.03	0.0117	0.32	0.03	0.0096
0.39	1	0.39	0.32	1	0.32
0.39	0.24	0.0936	0.32	0.24	0.0768
0.39	3.21	1.2519	0.32	3.21	1.0272
0.39	0.13	0.0507	0.32	0.13	0.0416
0.39	0.39	0.1521	0.32	0.39	0.1248
0.39	0.11	0.0429	0.32	0.11	0.0352
0.39	0.28	0.1092	0.32	0.28	0.0896
0.39	1.14	0.4446	0.32	1.14	0.3648
0.39	0.3	0.117	0.32	0.3	0.096
0.39	1.23	0.4797	0.32	1.23	0.3936
0.39	0.51	0.1989	0.32	0.51	0.1632
0.39	2.02	0.7878	0.32	2.02	0.6464
0.39	1.26	0.4914	0.32	1.26	0.4032
0.39	1.25	0.4875	0.32	1.25	0.4
0.39	1.28	0.4992	0.32	1.28	0.4096
0.39	0.59	0.2301	0.32	0.59	0.1888
0.39	1.42	0.5538	0.32	1.42	0.4544
0.39	1.27	0.4953	0.32	1.27	0.4064
0.39	0.22	0.0858	0.32	0.22	0.0704
0.39	0.15	0.0585	0.32	0.15	0.048
0.39	1.44	0.5616	0.32	1.44	0.4608
0.39	1.75	0.6825	0.32	1.75	0.56
0.39	1.08	0.4212	0.32	1.08	0.3456
0.39	2	0.78	0.32	2	0.64
0.39	1.03	0.4017	0.32	1.03	0.3296
0.39	0.89	0.3471	0.32	0.89	0.2848
0.39	0.47	0.1833	0.32	0.47	0.1504
0.39	1.67	0.6513	0.32	1.67	0.5344

0.39	0.18	0.0702	0.32	0.18	0.0576
0.39	2.54	0.9906	0.32	2.54	0.8128
0.39	0.1	0.039	0.32	0.1	0.032
0.39	0.44	0.1716	0.32	0.44	0.1408
0.39	0.41	0.1599	0.32	0.41	0.1312
0.39	1.45	0.5655	0.32	1.45	0.464
0.39	0.55	0.2145	0.32	0.55	0.176
0.39	1.5	0.585	0.32	1.5	0.48
0.39	0.29	0.1131	0.32	0.29	0.0928
0.39	0.3	0.117	0.32	0.3	0.096
0.39	1.79	0.6981	0.32	1.79	0.5728
0.39	0.88	0.3432	0.32	0.88	0.2816
0.39	0.54	0.2106	0.32	0.54	0.1728
0.39	1.06	0.4134	0.32	1.06	0.3392
0.39	0.79	0.3081	0.32	0.79	0.2528
0.39	1.08	0.4212	0.32	1.08	0.3456
0.39	0.2	0.078	0.32	0.2	0.064
0.39	0.1	0.039	0.32	0.1	0.032
0.39	0.76	0.2964	0.32	0.76	0.2432
0.39	1.98	0.7722	0.32	1.98	0.6336
0.39	1.05	0.4095	0.32	1.05	0.336
0.39	0.15	0.0585	0.32	0.15	0.048
0.39	0.28	0.1092	0.32	0.28	0.0896
0.39	1.11	0.4329	0.32	1.11	0.3552
0.39	0.07	0.0273	0.32	0.07	0.0224
0.39	2.09	0.8151	0.32	2.09	0.6688
0.39	1.55	0.6045	0.32	1.55	0.496
0.39	0.66	0.2574	0.32	0.66	0.2112
0.39	0.47	0.1833	0.32	0.47	0.1504
0.39	0.34	0.1326	0.32	0.34	0.1088
0.39	0.67	0.2613	0.32	0.67	0.2144
0.39	0.33	0.1287	0.32	0.33	0.1056
0.39	1.29	0.5031	0.32	1.29	0.4128
0.39	0.6	0.234	0.32	0.6	0.192
0.39	0.73	0.2847	0.32	0.73	0.2336
0.39	1.42	0.5538	0.32	1.42	0.4544
0.39	0.65	0.2535	0.32	0.65	0.208
0.39	0.15	0.0585	0.32	0.15	0.048
0.39	0.76	0.2964	0.32	0.76	0.2432
0.39	1.23	0.4797	0.32	1.23	0.3936
0.39	0.23	0.0897	0.32	0.23	0.0736

0.39	3.9	1.521	0.32	3.9	1.248
0.39	2.89	1.1271	0.32	2.89	0.9248
0.39	0.19	0.0741	0.32	0.19	0.0608
0.39	0.72	0.2808	0.32	0.72	0.2304
0.39	0.69	0.2691	0.32	0.69	0.2208
0.39	1.4	0.546	0.32	1.4	0.448
0.39	0.73	0.2847	0.32	0.73	0.2336
0.39	1.3	0.507	0.32	1.3	0.416
0.39	0.51	0.1989	0.32	0.51	0.1632
0.39	1.58	0.6162	0.32	1.58	0.5056
0.39	1.32	0.5148	0.32	1.32	0.4224
0.39	1.68	0.6552	0.32	1.68	0.5376
0.39	0.8	0.312	0.32	0.8	0.256
0.39	0.95	0.3705	0.32	0.95	0.304
0.39	0.87	0.3393	0.32	0.87	0.2784
0.39	0.45	0.1755	0.32	0.45	0.144
0.39	0.46	0.1794	0.32	0.46	0.1472
0.39	2.22	0.8658	0.32	2.22	0.7104
0.39	0.35	0.1365	0.32	0.35	0.112
0.39	1.74	0.6786	0.32	1.74	0.5568
0.39	1.26	0.4914	0.32	1.26	0.4032
0.39	1.94	0.7566	0.32	1.94	0.6208
0.39	1.63	0.6357	0.32	1.63	0.5216
0.39	0.13	0.0507	0.32	0.13	0.0416
0.39	0.45	0.1755	0.32	0.45	0.144
0.39	0.03	0.0117	0.32	0.03	0.0096
0.39	2.85	1.1115	0.32	2.85	0.912
0.39	1.36	0.5304	0.32	1.36	0.4352
0.39	0.76	0.2964	0.32	0.76	0.2432
0.39	1.25	0.4875	0.32	1.25	0.4
0.39	0.3	0.117	0.32	0.3	0.096
0.39	1.62	0.6318	0.32	1.62	0.5184
0.39	1.65	0.6435	0.32	1.65	0.528
0.39	0.4	0.156	0.32	0.4	0.128
0.39	1.25	0.4875	0.32	1.25	0.4
0.39	2.01	0.7839	0.32	2.01	0.6432
0.39	2.92	1.1388	0.32	2.92	0.9344
0.39	1.55	0.6045	0.32	1.55	0.496
0.39	0.08	0.0312	0.32	0.08	0.0256
0.39	1.53	0.5967	0.32	1.53	0.4896
0.39	2.39	0.9321	0.32	2.39	0.7648

0.39	0.82	0.3198	0.32	0.82	0.2624
0.39	2.18	0.8502	0.32	2.18	0.6976
0.39	0.49	0.1911	0.32	0.49	0.1568
0.39	0.13	0.0507	0.32	0.13	0.0416
0.39	1.47	0.5733	0.32	1.47	0.4704
0.39	1.14	0.4446	0.32	1.14	0.3648
0.39	0.78	0.3042	0.32	0.78	0.2496
0.39	0.47	0.1833	0.32	0.47	0.1504
0.39	2.72	1.0608	0.32	2.72	0.8704
0.39	1.33	0.5187	0.32	1.33	0.4256
0.39	1.66	0.6474	0.32	1.66	0.5312
0.39	0.89	0.3471	0.32	0.89	0.2848
0.39	0.44	0.1716	0.32	0.44	0.1408
0.39	0.45	0.1755	0.32	0.45	0.144
0.39	0.75	0.2925	0.32	0.75	0.24
0.39	0.9	0.351	0.32	0.9	0.288
0.39	0.53	0.2067	0.32	0.53	0.1696

Conclusion:

In the classical analysis approach of the field experimental data, the correlation present in the data is ignored. In the presence of significantly correlated trends in the data, we cannot neglect the correlation effect. Because of this reason it is therefore, becomes necessary to look for these situations and methods of analysis. Since theories for estimating the variance components are available in the literature only for uncorrelated errors, the theory for estimating the variance components in case of correlated structure will be developed. The influence of the correlated errors on the estimates of random effects is empirically examined. Again, using the concept of variance component estimates, further the effect of correlated errors as well as uncorrelated errors is checked. The estimate of important genetic parameters are empirically examined.

Objective-I

```
# The "diff-REML" function is used for robust DF-REML method
# The output includes:
#  -Coef: model coefficients (for intercept, PCs and SNPs)
#  -weights.sum: sum of the per observation weights used in the robust fit
#  -rand.var: estimated variance for the random effects (genetic variance)
#  -res.var: estimated variance for the residual errors
#  - Conditional.R2: Adjusted conditional R^2
#  - Marginal.R2: Ajusted marginal R^2

diff.REML<-function(response,fixed,random,method,intercept,ubound){
  ##
  ## response : continuous response variable; coded as numeric
  ##
  ## fixed    : matrix or dataframe of fixed effects == explanatory variables
  ##          coded appropriately
  ## random   : variance-covariance structure of the random effects
  ##          needs to be a symmetric positive semi-definite matrix
  ##          if it is a Kinship matrix K, multiply by two when calling the function function
  ##          if it is NRM, GRM or just the indentity call the function directly
  ##          throughtout the algorithm semi-definite matrices are needed; if problems
  ##          are encountered please refer to the paper of
  ##             Rousseeuw&Molenberghs (1993), Comm in Stats -Theory &Meths, 22(4), 965-984
  ## method   : DF-REML approach of the lmekin() function
  ##          from package coxme which optimizes a different likelihood
  ## intercept: one may choose between estimating or not the intercept
  ## ubound   : upper bound for the initial estimate of theta needed for the optimize() function
  ##          function instead of the optim() function, in which case one should change the code

require(MASS)        # for hubers() and rlm() functions
require(robustbase)  # for Qn() function

  y<-response; X<-fixed; K<-random
K<-as.matrix(K)
X<-as.matrix(X)
nn<- dim(X)[1]
pp<- dim(X)[2]

robfs<- function(xr){hubers(xr,mu=0)$s^2}
robfs<- function(xr){Qn(xr)^2}
robfs<- function(xr){mean(xr^2)}
```

```
# Step 2
  if(intercept==T){Xstar<- cbind(rep(1,nn),X)}else{Xstar<- X }

  # Qr decomposition of XX'; U matrix corresponds to the U' in the algorithm
mX<-Xstar%*%t(Xstar)

  # code added for the specific data set
eigenmX<-eigen(mX)$values
if(length(eigenmX[eigenmX<0])>0){
lambda<-abs(eigenmX[length(eigenmX)])+1e-9
mXpd<-(mX+lambda*diag(dim(mX)[1]))/(1+lambda)
rownames(mXpd)<-rownames(K)
colnames(mXpd)<-rownames(K)
mX<-mXpd
  }
  U<- t(qr.Q(qr(mX))[,((pp+1)+1):nn])

  # transforming phenotypes y
  y.dot<-U%*%y

# Steps 3-5
ploglik<- function(theta){                  # defining the profile log-likelihood function

    # defining ther variables as in the algorithm
phi<-theta*K+diag(nn)
    phi.dot<-(U%*%(phi)%*%t(U))
    Delta<- t(solve(chol(phi.dot)))

    # transforming phenotypes y.dot
y.star<-Delta%*%y.dot

    #profile log-likelihood
if(intercept==T){
loglik<- (nn-(pp+1))*(log(2*pi)+1+log(robfs(y.star)))+log(det(phi.dot))
}else{loglik<- (nn-pp)*(log(2*pi)+1+log(robfs(y.star)))+log(det(phi.dot))}
   }

# Step 6

  ## using the optimize function()
sol<- optimize(function(x) ploglik(x),c(0,ubound))     # optimizing the profile log-likelihood
function
newtheta<-sol$minimum

# Step 7
```

```
phi.dot.hat<- U%*%(newtheta*K+diag(nn))%*%t(U)
Delta.hat<- t(solve(chol(phi.dot.hat)))

  s    <- robfs(Delta.hat%*%y.dot)
if(intercept==T){
logl<- -(nn-(pp+1)
}else{logl<- -(nn-pp)*(log(2*pi)+1+log(s))/2-log(det(phi.dot.hat))/2}

# Step 8
  sa2r<- s*newtheta         # variance for the random effects (genetic variance)
  se2r<- s                  # variance for the residual errors

# Step 9
V.hat<-s*(newtheta*K+diag(nn))
  Delta.dot<- t(solve(chol(V.hat)))

# Step 10
fit<-rlm(Delta.dot%*%y~0+Delta.dot%*%Xstar)
     #use rlm(Delta.dot%*%y~0+Delta.dot%*%Xstar
#method="MM") in the case the predictor variables are continuous and may be contaminated
fit<-lm(Delta.dot%*%y~0+Delta.dot%*%Xstar)}
fit.coef<- fit$coef

# computing the marginal residuals
  y.res   <- y-Xstar%*%fit.coef

# computing the conditional residuals
rm(fit)
new.Delta<-  t(solve(chol(sa2r*K)))
if(method!="ls"){
fit<-  rlm(y.res~0+new.Delta)
       #use rlm(Delta.dot%*%y~0+Delta.dot%*%Xstar, method="MM") in the case the predictor
variables are continuous and may be contaminated
}else{fit <-  lm(y.res~0+new.Delta)}
rcoef<- fit$coef
y.hat<- Xstar%*%fit.coef+diag(length(rcoef))%*%rcoef
  fit.res   <- y-y.hat
rm(fit)

# Preparing the output
wsum<- sum(rlm(Delta.dot%*%y~0+Delta.dot%*%Xstar)$w)       # Sum of the per observation
weights used in the robust fit

fit.aux<- matrix(rep("NA",length(fit.coef)), ncol=1)
  if(intercept==T){row.names(fit.aux)<-              c("Intercept",              paste("X",1:dim(X)[2],
sep=""))}else{row.names(fit.aux)<-c(paste("X",1:dim(X)[2], sep=""))}
```

```
colnames(fit.aux)<- c("Estimate")
fit.aux<- as.data.frame(fit.aux)
fit.aux[,1]<- round(fit.coef, 5)

results<- list(Coef=fit.aux, rCoef=rcoef,weights.sum=wsum, rand.var=sa2r, res.var=se2r,
Conditional.R2=R2c, Marginal.R2=R2m,
Fixed.R2=sf2r,MargResiduals=y.res,CondResiduals=fit.res,loglik=logl)
return(results)

  }
```

Objective-II:

```
prociml;
vs=0.225; ve=0.775;
print vs ve;
ht=4*vs/(vs+ve);
printht;
p=1;
m=12;
rc1=1;
rc2=-2;
doiter=1to p;
doi=1to10;
si=sqrt(vs)*rannor(800);
do k=1to10;
if k=1then do;

eijk1=sqrt(ve)* rancau (12345)*rc1;
eijk2=sqrt(ve)* rancau (23145)*rc2;
end;
else do;
eijk1=rc1*eijk1+ rancau (1000);
eijk2=rc2*eijk2+ rancau (1000);
end;
z=m+si+eijk1+eijk2;
yd=yd//z;
sr1=sr1//i;
num=num//iter;
gm=gm//1;
end;
end;
end;
/*create no_selectvar{yd sr1 num gm};*/
/*append ; */
run;
print sr1 yd;
quitiml;


prociml;
*vs=0.225;
ve=1;
*print vs ve;
*ht=4*vs/(vs+ve);
*printht;
ht=0.2;
```

```
vs=ht/(4-ht);
print vs;
p=200;
m=12;
doiter=1to p;
doi=1to10;
si=sqrt(vs)*rannor(800);
do k=1to10;
if k=1then do;

eijk1=sqrt(ve)* rancau (12345)*rc1;
eijk2=sqrt(ve)* rancau (23145)*rc2;
end;
else do;
eijk1=rc1*eijk1+ rancau (1000);
eijk2=rc2*eijk2+ rancau (1000);
z=m+si+eijk;
yd=yd//z;sr1=sr1//i;num=num//iter;gm=gm//1;
end;
end;
end;
createno_selectvar{yd sr1 num gm};
append ;
run;
quitiml;
data pop1;
setno_select;
procmixedmethod=type3 noclprintnoinfonoitprint;
class sr1;
model yd=gm;
random sr1;
make'covparms' out=canv;
by num notsorted;
run;
data pop2;
setno_select;
procmixedmethod=ml noclprintnoinfonoitprint;
class sr1;
model yd=gm;
random sr1;
make'covparms' out=cml;
by num notsorted;
run;
data pop3;
setno_select;
procmixedmethod=remlnoclprintnoinfonoitprintnoprofile ;
class sr1;
```

```
model yd=gm;
random sr1;
make'covparms' out=creml;
by num notsorted;
run;
data pop4;
setno_select;
procmixedmethod=mivque0 noclprintnoinfonoitprintnoprofile ;
class sr1;
model yd=gm;
random sr1;
make'covparms' out=cmivque;
by num notsorted;
run;
prociml;
usecanv;
read all intoanvc;
p=200;v={0.0251};ht=4*v[1,1]/v[,+];
anvc=anvc[,2];
anvc=shape(anvc,p);
anvcmn=anvc[+,]/p;
printanvc;
anvcbias=anvcmn-v;
anht=4*anvc[,1]/anvc[,+];
anhtbias=anht-ht;
anhtbiasmn=anhtbias[+,]/p;
anhtmn=anht[+,]/p;
anhtbs=anhtmn-ht;
anhtmse=ssq(anht-ht)/p;
**print anvcmnanhtmnanhtbsanhtmse;
use cml;
read all intomlvc;
mlvc=mlvc[,2];
mlvc=shape(mlvc,p);
mlvcmn=mlvc[+,]/p;
mlvcbias=mlvcmn-v;
mlht=4*mlvc[,1]/mlvc[,+];
mlhtbias=mlht-ht;
mlhtbiasmn=mlhtbias[+,]/p;
mlhtmn=mlht[+,]/p;
mlhtbs=mlhtmn-ht;
mlhtmse=ssq(mlht-ht)/p;
**print mlvcmnmlhtmnmlhtbsmlhtmse;
usecreml;
read all intoremlvc;
remlvc=remlvc[,2];
remlvc=shape(remlvc,p);
```

```
remlvcmn=remlvc[+,]/p;
remlvcbs=remlvcmn-v;
remlht=4*remlvc[,1]/remlvc[,+];
remlhtbias=remlht-ht;
remlhtbiasmn=remlhtbias[+,]/p;
remlhtmn=remlht[+,]/p;
remlhtbs=remlhtmn-ht;
rmlhtmse=ssq(remlht-ht)/p;
**print remlvcmnremlhtmnremlhtbsrmlhtmse;
usecmivque;
read all intomivqvc;
mivqvc=mivqvc[,2];
mivqvc=shape(mivqvc,p);
mqvcmn=mivqvc[+,]/p;
mqvcbias=mqvcmn-v;
mivqht=4*mivqvc[,1]/mivqvc[,+];
mivqhtbias=mivqht-ht;
mivqhtbiasmn=mivqhtbias[+,]/p;
mivqhtmn=mivqht[+,]/p;
mivqhtbs=mivqhtmn-ht;
mvqhtmse=ssq(mivqht-ht)/p;
**print mqvcmnmivqhtmnmivqhtbsmvqhtmse;
vc=anvcmn//mlvcmn//remlvcmn//mqvcmn;
ht=anhtmn//mlhtmn//remlhtmn//mivqhtmn;
htbs=anhtbs//mlhtbs//remlhtbs//mivqhtbs;
mse=anhtmse//mlhtmse//rmlhtmse//mvqhtmse;
printvchthtbsmse;
run;
```

Objective-III:

```
# Data
#--------------------------------------#

# given parameters (page 37)
sigmaE<- 40
sigmaA<- 20
alpha  <- sigmaE / sigmaA
print(alpha)
# Data (Paul)
calf<- c(4:8)
sex<- c("male", "female", "female", "male", "male")
WWG    <- c(4.5, 2.9, 3.9, 3.5, 5.0)

# Dataframe
```

```
data.3.1 <- data.frame(calf, sex, WWG)
rm(list=(c("calf", "sex", "WWG")))

# print data
print(data.3.1)
#---------------------------------------#
# Pedigree
#---------------------------------------#

# set up pedigree
Calf    <- c(1:8)
Sire    <- c(0, 0, 0, 1, 3, 1, 4, 3)
Dam     <- c(0, 0, 0, 0, 2, 2, 5, 6)
ped.3.1 <- data.frame(Calf, Sire, Dam)
rm(list=c("Calf","Sire","Dam"))
print(ped.3.1)

# Create A ###############

`createA` <-

function(ped){

if (nargs() > 1 ) {
stop("Only the pedigree is required (Animal, Sire, Dam)")
   }

   # This is changed from Gota's function
   # Extract the sire and dam vectors
   s = ped[, 2]
   d = ped[, 3]

   # Stop if they are different lengths
if (length(s) != length(d)){
stop("size of the sire vector and dam vector are different!")
   }

   # set number of animals and empty vector
   n <- length(s)
   N <- n + 1
   A <- matrix(0, ncol=N, nrow=N)

   # set sires and dams
   s <- (s == 0)*(N) + s
   d <- (d == 0)*N + d
```

```
start_time<- Sys.time()
   # Begin for loop
for(i in 1:n){

     # equation for diagonals
A[i,i] <- 1 + A[s[i], d[i]]/2

for(j in (i+1):n){    # only do half of the matrix (symmetric)
if (j > n) break
A[i,j] <- ( A[i, s[j]] + A[i, d[j]] ) / 2  # half relationship to parents
A[j,i] <- A[i,j]    # symmetric matrix, so copy to other off-diag
     }
   }

   # print the time it took to complete
cat("\t", sprintf("%-30s:%f", "Time it took (sec)", as.numeric(Sys.time() - start_time)), "\n")

   # return the A matrix
return(A[1:n, 1:n])

  }

############################
library("Matrix")
A <- Matrix(createA(ped.3.1))
A <- as(A, "sparseMatrix")
print(A)

# inverse of A
Ainv<- solve(A)
print(Ainv)

# set response variable
y <- Matrix(data.3.1$WWG, ncol=1)
print(y)
data.3.1$sex <- factor(data.3.1$sex, levels=c("male", "female"))

# set up X
X <- Matrix(model.matrix(~ sex - 1, data=data.3.1))
print(X)

# set up Z (n by n*) n* = number of animals in whole pedigree (8)
Z <- Matrix(0, nrow=nrow(data.3.1), ncol=nrow(ped.3.1))

# change row/column names
colnames(Z) <- ped.3.1$Calf
```

```
rownames(Z) <- data.3.1$calf

# fill in Z
for (i in 1:nrow(data.3.1)) {
index = data.3.1[i, "calf" ]
Z[i, index] = 1
}

# print Z
print(Z)

# set up MME for PE solutions
basicMME<- function(X, Z, A, y, alpha) {

  # Calculate blocks of LHS
XpX = crossprod(X)                    # same as t(X) %*% X
ZpZ = crossprod(Z) + (solve(A) * alpha)      # same as (t(Z) %*% Z) + (solve(A)*alpha)
XpZ = crossprod(X, Z)                  # same as t(X) %*% Z

  # Paste top and bottom together for LHS
toprow    = cbind(XpX,    XpZ)
bottomrow = cbind(t(XpZ), ZpZ)

  # Put top and bottom together for left hand side (LHS)
  LHS = rbind(toprow, bottomrow)

  # Elements for RHS
XpY = crossprod(X, y)
ZpY = crossprod(Z, y)

  # Calculate right hand side (RHS)
  RHS = rbind(XpY, ZpY)

  # calculate solutions by direct inversion of LHS
solutions = solve(LHS) %*% RHS

  # Return LHS, RHS, and solutions
return(list(LHS=LHS, RHS=RHS, solutions=solutions))

}

output<- basicMME(X, Z, A, y, alpha)
# print the LHS
LHS <- output[[1]]
print(LHS)
```

```
# print the inverse of LHS
LHSinv<- solve(LHS)
print(round(LHSinv, 4))

# print the RHS
RHS <- output[[2]]
print(RHS)

# print the solutions
solutions<- output[[3]]
rownames(solutions) <- rownames(RHS)
print(solutions)

# variance of a
G <- A*sigmaA

# print G
print(G)

# variance of e
R <- Matrix(diag(nrow(data.3.1))*sigmaE)

# print R
print(R)

# variance of y
V <- Z %*% G %*% t(Z) + R

# print V
print(V)

# Take the inverse of the LHS and multiply by sigma^2_e
solve(LHS)
solve(LHS)*sigmaE

# create data frame
data.acc<- data.frame(Animal = 1:8)

# create column with diagonals
data.acc$DiagLHS = diag(solve(LHS))[3:10]

# calculate relability from Diagonals of LHS (alpha = sigmaE / sigmaA)
data.acc$r2 = 1 - data.acc$DiagLHS*alpha

# calculate the accuracy from reliability (r^2)
data.acc$r = sqrt(data.acc$r2)
```

```
# calculate Standard Error of Prediction (SEP) (using diagonals and the variance of e)
data.acc$SEP = sqrt(data.acc$DiagLHS*sigmaE)

# now the variance (square the standard error of prediction (SEP))
data.acc$PEV = data.acc$SEP^2

# print
print(data.acc)


############# NOT REQUIRED##############
########################################

# histograms
data.acc %>%
gather(Type, Value, DiagLHS:PEV) %>%
  mutate(Type = fct_relevel(Type, "DiagLHS", "r2", "r", "SEP", "PEV")) %>%
ggplot(., aes(x=Value)) +
geom_histogram(bins=5, color="white", fill="dodgerblue3") +
facet_wrap(~ Type, scales="free") +
theme_bw()
#=================================================================
==========#
# Plot BIF vs True accuracy
#=================================================================
==========#

# BIF accuracy = 1 - sqrt(1 - r^2)

# Accuracy = sqrt(1 - (1-BIF)^2)

#=================================================================
==========#
# Setup
#=================================================================
==========#

# my ggplot2 theme
my_gg_theme<- theme_light() +
theme(text = element_text(size=16),
plot.title = element_text(hjust=0.5),
panel.grid = element_blank())

#=================================================================
==========#
```

```
# Code
#=================================================================
==========#

# vector of accuracies
Acc<- seq(0,1,by=0.01)
r2     <- Acc^2
BIFacc<- 1 - sqrt(1 - Acc^2)

# combine into dataset
data<- data.frame(Seq = seq(0,1,by=0.01), Acc = Acc, r2 = r2, BIFAcc = BIFacc)

# melt
data.gather<- gather(data, Type, Value, Acc:BIFAcc) %>%
mutate(Type = fct_relevel(Type, "Acc", "r2", "BIFAcc"))

# plot
print(ggplot(data.gather, aes(x=Seq, y=Value, color=Type, shape=Type, group=Type)) +
geom_line() +
geom_point() +
scale_color_manual("Accuracy", values=c("dodgerblue3", "blue", "magenta"),
labels = c("True Acc", "Reliability", "BIF Acc")) +
scale_shape_manual("Accuracy", values=c(16,17,15),
labels = c("True Acc", "Reliability", "BIF Acc")) +
ylab("Reported Accuracy") +
xlab("True Accuracy") +
my_gg_theme)

#Illustration through simulation
# load MASS for mvrnorm function
library(MASS)
# we can generate TBVs using the mean (0) and variance (A*sigma2_a)
mu = rep(0, 8)
Sigma = A*sigmaA

# create 1000 samples of TBVs (random effects change each sampling)
TBV.samples<- mvrnorm(n=10000, mu=mu, Sigma=Sigma)

# row/col names
rownames(TBV.samples) <- 1:nrow(TBV.samples)
colnames(TBV.samples) <- 1:8

# now look at the correlations between them!
round(cov(TBV.samples), 3)
# multiply A and sigma^2_a
A * sigmaA
```

```
# now look at the correlations between them!
round(cor(TBV.samples), 3)
# sex
sex<- c("male", "female", "female", "male", "male")

# Let us assume that the TRUE fixed effects are really 4.4 and 3.4
sex.solutions<- Matrix(c(4.4, 3.4), ncol=1, nrow=2)

# Get fixed part of Model
fixed<- X %*% sex.solutions

# replicate
fixed<- Matrix(rep(fixed, 10000), nrow=5, ncol=10000)

# Get random part of Model (Z is (5,8) and TBV is (10000,8)
# so the result will be (5,10000)
random_a<- Z %*% t(TBV.samples)

# so to generate y we also need the residuals
random_e<- t(mvrnorm(n=10000, mu=rep(0,5), Sigma=diag(5)*sigmaE))

# finally we can generate y
random_y<- data.frame(as.matrix(t(fixed + random_a + random_e)))

# cov
cov(random_y)
# variance for y
Z%*%G%*%t(Z) + R
# mean of male (animals 1, 4, 5) and female (animals 2 and 3).
mean(random_y[, 1])
mean(random_y[, 4])
mean(random_y[, 5])
# females
mean(random_y[, 2])
mean(random_y[, 3])
```

सारांश और निष्कर्ष

पौधे और पशु प्रजनन में कई प्रयोगात्मक स्थितियों में अवलोकन स्वतंत्र नहीं हैं तथा अवलोकनों के बीच किसी प्रकार का सहसंबंध मौजूद है। डेटा में महत्वपूर्ण सहसंबद्ध प्रवृत्ति की उपस्थिति में, स्वतंत्रता की पारम्परिक धारणा का उल्लंघन होता है। महत्वपूर्ण लक्षणों की प्रसरण के आनुवंशिक घटकों की जानकारी पौधे और पशु प्रजनकों की प्रमुख रुचि है। इस प्रकार, विभिन्न आनुवंशिक भिन्नताओं का अनुमान और उनकी वंशानुक्रम के बारे में अनुमान, विभिन्न आनुवंशिक मापदंडों के अनुमानों के आधार पर पौधे और पशु प्रजनन कार्यक्रम के दृष्टिकोण से बहुत महत्वपूर्ण है।

इसके अलावा मिश्रित मॉडलों में, स्थिर प्रभावों की तुलना के अलावा, पौधे और पशु प्रजनकों की मुख्य रुचि महत्वपूर्ण लक्षणों के प्रसरण के आनुवंशिक घटकों के बारे में जानकारी रखना है। इसके परिणामस्वरूप विभिन्न आनुवंशिक भिन्नताओं का अनुमान लगाया जाता है और विभिन्न आनुवंशिक मापदंडों के अनुमानों के आधार पर उनकी वंशानुक्रम के बारे में अनुमान लगाया जाता है। साहित्य में, उन मॉडलों के लिए प्रसरण घटकों का अनुमान लगाने के लिए सिद्धांत उपलब्ध है जिनमें असहसंबद्ध त्रुटियां हैं लेकिन सहसंबद्ध संरचना के लिए नहीं। इसलिए, सहसंबद्ध त्रुटि संरचना के मामले में प्रसरण घटकों के आकलन के लिए सिद्धांत विकसित करना आवश्यक था ताकि प्रयोगात्मक स्थितियों की आवश्यकताओं को पूरा किया जा सके। फिर, जटिल त्रुटि संरचना में, जिसे मॉडल नहीं किया जा सकता है, कोई भी प्रयोगसिद्ध अध्ययनों के माध्यम से प्रसरण घटकों के अनुमान पर सहसंबंध के प्रभाव का अध्ययन कर सकता है। इसलिए, त्रुटियों के सहसंबद्ध होने पर आनुवंशिक मापदंडों के आकलन के लिए सांख्यिकीय दृष्टिकोण विकसित करना सांख्यिकीय आनुवंशिकी के क्षेत्र में एक महत्वपूर्ण शोध योग्य क्षेत्र है।

इस अध्ययन में, हमने एसएनपी–आधारित आनुवांशिकता का अनुमान लगाने के लिए इस अध्ययन में रोबस्ट आनुवांशिकता आंकलन पद्धतियों का प्रयोग किया, जब डेटासेट में अवलोकन को आउटलिंग अवलोकन से बिगाड़ा गया था। एलएमएम का उपयोग करते हुए प्रसरण घटकों का आकलन करने के लिए, रोबस्ट डीएफ–आरईएमएल आकलन पद्धति का वर्णन किया गया है। आंकलन तकनीक और निर्धारण के गुणांकों की व्युत्पत्ति के लिए, ह्यूबर (1964) और रूसो और क्राउक्स (1993) के पैमाने के रोबस्ट आकलनों का उपयोग किया गया था। वास्तविक डेटासेट पर रोबस्ट दृष्टिकोण लागू किए गए थे, और कई प्रक्रियाओं का पालन किया गया था, जैसे कि मूल डेटा और अशुद्ध डेटा (1%, 3% और 5%) दोनों का उपयोग करके आनुवांशिकता और प्रसरण घटकों का अनुमान लगाना। अंत में रोबस्ट दृष्टिकोण की तुलना पारम्परिक दृष्टिकोण से की जाती है। यह देखा गया है कि रोबस्ट आंकलनकर्ताओं ने पारम्परिक दृष्टिकोण से बेहतर प्रदर्शन किया।

वर्तमान अध्ययन में हमने डेटा जनरेशन के लिए हाफ सिब और फुल सिब मॉडल दोनों पर विचार किया। यहां हमने एआर (1) और एआर (2) दोनों त्रुटियों का इस्तेमाल किया। पूर्ण सिब मॉडल के मामले में हमने त्रुटि में एआर (1) सहसंबंध की उपस्थिति में ई (एमएसई) और आनुवांशिकता के आंकलन के लिए सूत्र निकाले। विभिन्न आनुवंशिकता मूल्यों के लिए 0.1,,0.5 को अलग–अलग एआर (1) मान -1 से +1 तक विचार किया है। जब सहसंबंध ऋणात्मक होता है तो त्रुटि के कारण वर्गों के योग के अपेक्षित माध्य को कम करके आंका जाता है और सहसंबंध की डिग्री बढ़ने पर वे बढ़ जाते हैं। लेकिन इन अपेक्षित माध्य वर्ग के योग को कम करके आंका जाता है यदि त्रुटियों को सकारात्मक रूप से सहसंबद्ध किया जाता है और वे सहसंबंध की डिग्री में वृद्धि के साथ घटते हैं और 0 तक पहुंच जाते हैं क्योंकि ρ ईकाई की ओर जाता है। दूसरी ओर, सायर के कारण वर्गों के योग का माध्य आंकलन करने के लिए केवल विपरीत परिणाम प्राप्त किए जाते हैं, अर्थात वर्गों के योग का अपेक्षित माध्य तब आंकलन होता है जब रो ऋणात्मक होता है और यदि सहसंबंध सकारात्मक होता है तो उन्हें कम करके आंका जाता है। जैसा कि अपने अधिकतम मूल्य के लिए सायर के दृष्टिकोण के कारण वर्गों के योग का अपेक्षित माध्य की ρ ईकाई की ओर जाता है। यदि सहसंबंध सकारात्मक है तो आनुवंशिकता मूल्यों का आंकलन लगाया जाता है। आनुवंशिकता के सभी स्तरों के लिए एक ही प्रवृत्ति का अनुसरण किया जाता है। इसके अलावा आनुवांशिकता शून्य से लगभग चार तक बढ़ जाती है क्योंकि ऑटोरेग्रेसिव गुणांक शून्य से ईकाई से आंकलन ईकाई तक बढ़ जाता है। डेटा ने विभिन्न आनुवंशिकता मूल्यों यानी उच्च, मध्यम और निम्न (0.5, 0.25, 0.1) हाफ सिब एआर (1) और विभिन्न नमूना आकार 100,200 और 500 और त्रुटियों के विभिन्न सहसंबंधों एआर (1) और एआर (2) का उपयोग करके उत्पन्न किया है। ρ=−1 से +1। डेटा प्रसरण घटकों को उत्पन्न करने के बाद SAS ProcVarcomp का उपयोग करके आंकलन लगाया जाता है। ANOVA, ML, REML और MIVQUE विधियों का उपयोग किया जाता है। सहसंबद्ध त्रुटियों एआर(1) और एआर (2) के मामले में आनुवांशिकता और एमएसई मूल्यों का हाफ सिब अनुमान और पूर्ण सिब आनुवांशिकता के पैरामीट्रिक मूल्य के लिए अलग–अलग नमूना आकार 0.10,0.25,0.5 विभिन्न नमूना आकारों के लिए 100, 200 और 500 के साथ और बिना किसी निश्चित प्रभाव के प्राप्त होते हैं। जब ρ −1 से +1 में बदलता है, तो आनुवंशिकता के अनुमानों का मान ऋणात्मक से धनात्मक में बदल जाता है। MSE मान ρ=0 तक घट रहा है, फिर बढ़ना ρ सकारात्मक है। ρ = −1 से ρ= −0.5 तक ML,REML और MIVQUE और MSE विधियों के मामले में मान नहीं बदल रहे हैं आनुवंशिकता का आंकलन 0 है। आनुवंशिकता का आंकलित मूल्य ρ=−.4 से ρ=1 तक बढ़ रहा है। MSE मान समान प्रवृत्ति दिखा रहे हैं। नमूना आकार बढ़ने से यह देखा गया है कि एमएसई मान घट रहे हैं।

AR(2) के मामले में यदि AR(1) मान को निर्धारित कर देते एआर (2) मूल्यों को बदलते है, तो एमएसई मूल्य सामान्य रूप से सहसंबंध मूल्य में वृद्धि के साथ कम हो जाता है। कभी–कभी क्रमरहित प्रवृत्ति देखी जाती हैं। हमने पाया कि AR(1) और AR(2) मानों (0, 0), (0.1,0.1) और (0.1, 0.5) के कुछ अच्छे संयोजन आनुवंशिकता का बेहतर आंकलन दे रहे हैं। नमूना आकार बढ़ने से यह देखा गया है कि एमएसई मान घट रहे हैं। लगभग सभी मामलों में पूर्वग्रह आंकलन प्राप्त किए जाते हैं। केवल सायर घटकों पर विचार करने का आंकलन केवल सायर और डैम घटक और डैम घटक पर विचार करने से बेहतर है। लगभग सभी मामलों में पूर्वग्रह आंकलन प्राप्त किए जाते हैं। केवल सायर घटकों पर विचार करने का आंकलन केवल सायर और डैम घटक और डैम घटक पर विचार करने से बेहतर है। एआर (2) के मामले में यदि एआर (1) 0.0 मान को निर्धारित करने पर एआर (2) मूल्यों को बदल रहा है, तो एमएसई मूल्य सामान्य रूप से सहसंबंध मूल्य में वृद्धि के साथ कम हो जाता है। कभी–कभी क्रमरहित प्रवृत्ति देखे जाते हैं। हमने पाया कि AR(1) और AR(2) मानों (–0. 5,–0.1) और (0.1,0.5) संयोजनों का कुछ अच्छा संयोजन आनुवांशिकता का बेहतर आंकलन दे रहा है। नमूना आकार बढ़ने से यह देखा गया है कि एमएसई मान घट रहे हैं। सभी मामलों के पूर्वग्रह आंकलन प्राप्त किए जाते हैं। फिक्सिंग के बाद एमएसई मान घटता है (एआर (1) मान एआर (2) मानों को एआर (1) मान 0.0 तक बढ़ाते हैं, यह प्रवृत्ति का अनुसरण करता है। सहसंबंध का संयोजन यानी (–.5, –.1) और (0.1, .0.5) किसी भी अन्य संयोजन की तुलना में बेहतर परिणाम दे रहे हैं। निश्चित प्रभाव के तहत भी सहसंबंध एआर (1) को –1 से +1 में 0.5 वृद्धि के साथ बदलने के बाद यह देखा गया है कि एआर (1) सहसंबंधों की वृद्धि के साथ एमएसई मान घटता है। सभी मामलों में प्राप्त आनुवंशिकता के अत्यधिक पूर्वग्रह आंकलन लगभग सभी मामलों में पूर्वग्रह आंकलन प्राप्त किए जाते हैं। केवल सायर घटकों पर विचार करने का अनुमान केवल सायर और डैम घटक और डैम घटक पर विचार करने से बेहतर है। सहसंबंध का संयोजन यानी (0,–0.5) और (0, 0.5) किसी भी अन्य संयोजन की तुलना में बेहतर परिणाम दे रहा है। नमूना आकार बढ़ाने से MSE मान कम हो जाते हैं।

हम यह निष्कर्ष निकाल सकते हैं कि विभिन्न सहसंबंध संरचनाएं अर्थात AR(1), AR(2) आदि का आनुवंशिकता के आंकलन और माध्य वर्ग त्रुटि मानों पर प्रभाव पड़ता है। बेहतर परिणामों के लिए अलग–अलग मामलों में अलग–अलग संयोजनों की पहचान की जाती है।

सायर घटक और त्रुटि घटक वितरण के विभिन्न संयोजनों के बाद उत्पन्न होते हैं जैसे नार्मल, बीटा, कॉची और टी–वितरण विभिन्न आनुवंशिकता मानों के साथ और आनुवंशिकता का आंकलन और चार अलग–अलग तरीकों से प्राप्त आरएमएसई मानों जैसे एनोवा, एमएल, आरईएमएल और एमआईवीक्यूई विधियों के साथ अलग–अलग पैरामीट्रिक आनुवंशिकता के मान है।

परिणामों से यह देखा गया है कि जब हम सहसंबंध मान ऋणात्मक को शून्य तक बढ़ाते हैं तो rmse मान कम हो जाते हैं। यदि हम शून्य से उच्चतर की ओर बढ़ते हैं, तो +1 के निकट यह देखा गया है कि वितरण के सभी संयोजनों के लिए आरएमएसई मानों में वृद्धि हुई है।

Summary and Conclusion

In many practical situations in the plant and animal breeding that observations are not independent and some kind of correlation among the observations exists. In the presence of significantly correlated trends in the data, the classical assumption of independence between observations is violated. The information on genetic components of variances of the important characters is the prime interest of plant and animal breeders. Thus, the estimation of various genetic variances and inferring about their inheritance, based on estimates of the different genetic parameters is very important from plant and animal breeding programme point of view.

Further in mixed models, in addition to comparison of fixed effects the prime interest of plant and animal breeders is to have information on the genetic components of the variances of the important characters. This results in estimating various genetic variances and inferring about their inheritance, based on estimates of the different genetic parameters. In literature, theory is available for estimating the variance components for those models which are having uncorrelated errors but not for correlated structure. So, it was required to develop the theory for estimating variance components in case of correlated error structure so as to meet the requirements of practical situations. Again, in the complicated error structure, which cannot be modelled, one can study the influence of correlation on the estimate of variance components through empirical studies. Hence, developing statistical approach for estimation of genetic parameters when errors are correlated is an important researchable area in the arena of statistical genetics.

In this study, we used robust heritability estimation methodologies in this study to estimate SNP-based heritability when observations in the dataset were contaminated with outlying observations. For estimating variance components using LMMs, the robust DF-REML estimate method has been described. For the estimation technique and the derivation of the coefficients of determination, Huber's (1964) and Rousseeuw and Croux's (1993) robust estimations of scale were used. The robust approaches were applied to a real dataset, and several procedures were followed, such as estimating heritability and variance components using both original and contaminated data (1%, 3% and 5%). Finally the robust approach is compared with classical

approach. It is seen that as contamination percentage increase, the estimated heritability decreases.

In the present study we considered both half sib and full sib model for data generation. Here we used both AR(1) and AR(2) errors. In case full sib model we derived formulae for E(MSE) and Estimate of heritability in the presence of AR(1) correlation in the error. For different heritability values 0.1,,0.5 are considered different AR(1) values from -1 to +1 are considered . The expected mean sum of squares due to error are overestimated when the correlation is negative and they increase as degree of correlation increase. But these expected mean sum of square are underestimated if errors are positively correlated and they decrease with increase in degree of correlation and approach to 0 as ρ tends to unity. On the other hand, just reverse results are obtained for estimating the mean sums of squares due to sire i.e. expected mean sum of squares are under estimated when rho is negative and they are overestimated if the correlation is positive. As ρ tends to unity expected mean sum of squares due to sire approaches to its maximum value.

Heritability values are over estimated if the correlation is positive. The same trend follows for all the level of heritability. Also heritability increases zero to nearly four as autoregressive coefficients increase from minus unity to approximate unity. Data have generated using different heritability values i.e. high, moderate and low(0.5, 0.25, 0.1) Half sib AR(!):and different sample size 100,200 and 500 and different correlation of errors A(1) and AR(2). ρ=-1 to +1. After generating the data variance components are estimated using SAS ProcVarcomp . ANOVA,ML,REML and MIVQUE methods are used. Half sib Estimate and full sib of heritability and MSE values in case of correlated errors (AR(1)) and AR(2) different sample sizes for parametric value of heritability 0.10,0.25,0.5 for different sample sizes 100, 200 and 500 with and with out fixed effects are obtained. Value of estimates of heritability changes negative to positive when ρ changes from -1 to +1. MSE value decreasing up to ρ=0, then again increasing ρ is positive. Up to ρ= -1 to ρ= -0.5 the estimate of heritability is 0 in case of ML,REML and MIVQUE methods and MSE values are not changing. Estimate value of heritability are increasing from ρ=-.4 to ρ=1 .MSE values are showing the same trend. Increasing sample sizes it is noticed that the MSE values are decreasing.

In case of AR(2) if fixing AR(1) value changing AR(2) values, the MSE value decrease with increasing the correlation value in general. Sometimes haphazard trends are noticed. We found some good combination of AR(1) and AR(2) values (0, 0) ,(0.1,0.1) and (0.1, 0.5) combinations are giving better estimates of heritability. Increasing sample sizes it is noticed that the MSE values are decreasing. In almost all cases biased estimates are obtained. Estimates for considering only sire components are better than considering sire and dam component and dam component alone. In almost all cases biased estimates are obtained. Estimates for considering only sire components are better than considering sire and dam component and dam component alone. In case of AR(2) if fixing AR(1) value changing AR(2) values, the MSE value decrease with increasing the correlation value in general. Sometimes haphazard trends are noticed. We found some good combination of AR(1) and AR(2) values (-0.5, -0.1)),) and (0.1, 0.5) combinations are giving better estimates of heritability. Increasing sample sizes it is noticed that the MSE values are decreasing. All the cases biased estimates are obtained. MSE values decreases after fixing (AR(1) values increase AR(2) values up to AR(1) values 0.0 it follows the trend. .The combination of correlation i.e. (-.5,-.1)) and (0.1, .0.5) are giving better results than any other combination. Under fixed effect also after changing correlation AR(1) from -1 to +1 with increment .5 it is noticed that MSE values decreases with increase of AR(1) correlations. All the cases highly biased estimates of heritability obtained. In almost all cases biased estimates are obtained. Estimates for considering only sire components are better than considering sire and dam component and dam component alone. The combination of correlation i.e. (0,-0.5) and (0, 0.5) are giving better results than any other combination. Increasing sample size decrease the MSE values.

We may conclude that different correlation structures i.e. AR(1), AR(2) etc. have impact on estimate of heritability and Means Square Error values. Different cases different combination are identified for better results. Sire component and error component are generated following different combination of distribution i.e. normal, beta, Cauchy and t-distribution with different heritability values and estimate of heritability and RMSE values obtained by four different methods i.e. ANOVA, ML,REML and MIVQUE methods with different parametric values of heritability.

From the results it is noticed that when we increase correlation value –ve to zero the rmse values decrease . If we increase from zero to higher i,e. nearer to +1 it is noticed that rmse values increased for all the combinations of distribution.

References

Akkaya, A. D., & Tiku, M. L. (2005). Robust estimation and hypothesis testing under short-tailedness and inliers. *Test*, **14**(1), 129-150.

Bernal-Vasquez, A. M., Möhring, J., Schmidt, M., Schönleben, M., Schön, C. C., & Piepho, H. P. (2014). The importance of phenotypic data analysis for genomic prediction-a case study comparing different spatial models in rye. *BMC genomics*, **15**(1), 1-17.

Bernal-Vasquez, A. M., Utz, H. F., & Piepho, H. P. (2016). Outlier detection methods for generalized lattices: a case study on the transition from ANOVA to REML. *Theoretical and Applied Genetics*, **129**(4), 787-804.

Cochran, W.G. (1939). The use of analysis of variance in enumeration by sampling. *J. Amer. Stat. Assoc.* **34**, 492-510.

Costa, C. N. , Carvalheira, J.A. , Cobuci,.A. F., Freitas,A.F. and Thompson,G. (2009) Estimation of genetic parameters of test day fat and protein yields in Brazilian Holstein cattle using an autoregressive multiple lactation animal model. *South African Journal of Animal Science*, 39:165-168.

Demidenko, E. (2013). *Mixed models: theory and applications with R*. John Wiley & Sons.

Diblasi, A. and Bowman, A.W. (2001). On the use of the variogram in checking for independence in spatial data. *Biometrics*, **57**, 211-218.

Durbin, J. and Watson, G.S. (1950). Testing for serial correlation in least squares regression. *Biometrika*, **37**, 409-428.

Fisher, R.A. (1925). *Statistical Methods for Research Workers*, 1st edn. Oliver & Boyd, Edinburgh and London.

Estaghvirou, S. B. O., Ogutu, J. O. & Piepho, H. P. (2014). Influence of outliers on accuracy and robustness of methods for genomic prediction in plant breeding. *G3*, **4**, 2317–2328.

Estaghvirou, S. B. O., Ogutu, J. O., Schulz-Streeck, T., Knaak, C., Ouzunova, M., Gordillo, A., & Piepho, H. P. (2013). Evaluation of approaches for estimating the accuracy of genomic prediction in plant breeding. *BMC genomics*, **14**(1), 1-21.

Fisher, R.A. (1925). *Statistical Methods for Research Workers*, 1st edn. Oliver & Boyd, Edinburgh and London.

Gilmour, A.R., Cullis, B.R. and Verbyla, A.P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *J. Agril., Bio.,and Env. Stats*, **2**, 269-293.

Graser, H. U., Smith, S. P., & Tier, B. (1987). A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *Journal of animal science*, **64**(5), 1362-1370.

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

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

Lourenço, V. M., Pires, A. M., & Kirst, M. (2011). Robust linear regression methods in association studies. *Bioinformatics*, **27**(6), 815-821.

Lourenço V M, Ogutu J O and Piepho H P(2020) Robust estimation of heritability and predictive accuracy in plant breeding:evaluation using simulation and empirical data, *BMC Genomics* 21-43.

Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibián-Barrera, M. (2019). *Robust statistics: theory and methods (with R)*. John Wiley & Sons.

Pinheiro, J., & Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

Renaud, O., & Victoria-Feser, M. P. (2010). A robust coefficient of determination for regression. *Journal of Statistical Planning and Inference*, **140**(7), 1852-1862.

Ronningen, K. (1974). Monte-Carlo simulation of statistical biological models which are of interest in animal breeding. *ActaAgriculturaeScandinavica*, **24**, 135-142.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American statistical association*, **79**(388), 871-880.

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

Singh, N. Okendro ,Bhatia V.K. and Paul, A.K. (2006) Estimation of Variance Components When Errors Are Correlated by Autoregressive of Order One ,The Journal of Indian Society of Agricultural Statistics, **60**(2):126-136

Wahi, S.D. and Rao, A.R.(2004) Some empirical investigations on the influence of fixed effects on the estimates of heritability. Unpublished Priject report, IASRI, New Delhi.

Estaghvirou, S. B. O., Ogutu, J. O. & Piepho, H. P. (2014). Influence of outliers on accuracy and robustness of methods for genomic prediction in plant breeding. *G3,* **4**, 2317–2328.

Fisher, R.A. (1925). *Statistical Methods for Research Workers*, 1st edn. Oliver & Boyd, Edinburgh and London.

Gilmour, A.R., Cullis, B.R. and Verbyla, A.P. (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *J. Agril., Bio.,and Env. Stats*, Vol.**2**, 269-293.

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

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

Paul A K and Wahi S D(2016) Estimation of heritability under correlated errors. Project Report. ICAR-IASRI publication.

Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., ... & Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics*, **38**(2), 203-208.

Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., ... & Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature genetics*, **42**(4), 355-360.

Zhao, K., Aranzana, M. J., Kim, S., Lister, C., Shindo, C., Tang, C., ... & Nordborg, M. (2007). An Arabidopsis example of association mapping in structured samples. *PLoS genetics*, **3**(1), e4.

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*, **109**(4), 1193-1198.