Crop Breeding and Applied Biotechnology 5:265-271, 2005
Brazilian Society of Plant Breeding. Printed in Brazil

Prediction of genetic value in F_3 populations of *Avena sativa* L. using Reml/Blup

Jefferson Luís Meirelles Coimbra[1*], Mauricio Marini Kopp[1], Velci Queiroz de Souza[1], Giovani Benin[2], Volmir Sergio Marchioro[3], Fernando Iraja Felix de Carvalho[1], and Antonio Costa de Oliveira[1]

Received 22 January 2005

Accepted 24 June 2005

ABSTRACT - *In genetics and breeding studies it is common to conduct experiments of fixed (sowing method) and random (populations) factors. Therefore, the most appropriate statistic analyses would use mixed linear models. In this sense, objectives of this work were i) to estimate the variance components for the random factor effects and interaction population x sowing method; ii) to perform selection among populations and among populations in each sowing method; iii) to compare the effect of the fixed factor. Both the effect of components of the population variance as well as the effect of the single interaction population x sowing methods revealed low contributions to the total variance. It can therefore be concluded that segregating populations present narrow genotypic variability. The Reml/Blup procedure is indicated for estimation/prediction in oat improvement where experiments generally produce unbalanced data.*

Key words: mixed linear models, linear prediction**,** sowing methods.

INTRODUCTION

One of the main contributions of quantitative genetics to plant breeding has been to make variance component estimates possible (Ramalho and Vencovsky 1978). According to Scheffé (1959) the mixed model was extensively reported by Fisher in 1918 in studies on covariance and genetic correlation between parents. He was the first to partition the genetic variance of an outcrossing population in three components: i) additive genetic variance, which is due to the mean gene effects; ii) dominant genetic variance, which is due to the intra-locus interaction effects; iii) epistatic genetic variance, based on the inter-locus interaction effects. Fisher also demonstrated that covariance between parents is a function of differences between variance components.

By concept, a variance component is variance associated to random effects of a model. Knowledge on this subject is essential for genetics and breeding (Littel et al. 1996). As a consequence, the development of more efficient selection methods for plant breeding depends on this kind of information, for example, to understand the genetic variability and the predominant gene action type controlling the trait under selection, which can be predicted

[1]Departamento de Fitotecnia, "Faculdade de Agronomia – Eliseu Maciel", Universidade Federal de Pelotas, C. P. 354, 96.001-970, Pelotas, RS, Brasil.
[*]E-mail: coimbrajefferson@pop.com.br
[2]Fundação Pró Sementes de Apoio a Pesquisa, Rua Diogo de Oliveira, 640, 99.025-130, Passo Fundo, RS, Brasil.
[3]Programa de trigo da Cooperativa Central de Pesquisa Agrícola (COODETEC), BR 476, km 98, 85.818-660, C. P. 301, Cascavel, PR, Brasil.

by means of variance components.

The effect of a factor can be defined as fixed or random. If a given factor is considered fixed, naturally, the goal will be to estimate and test hypotheses on linear combinations. However, in the case of considering it random, the aim is the estimation of variance components, once its levels are considered a random sample of a given study target genotype (Bueno Filho and Vencovsky 2000).

According to Henderson (1975), when considering allelic segregations in which each genotype is a vector of alleles that segregate and unite to form new genotypes, we have the situation in which selected individuals from each cross represent a possible progeny sample. However, in the case of assuming genetic effects as random, the methodology of variance components can be employed advantageously. Here, the random effects are predicted using the Best Linear Unbiased Predictor - Blup, which is the most appropriate method of prediction of genetic values, including the use for the prediction of virtual crosses (Bernardo 1995). Henderson (1986) postulated that the main restriction for the use of this methodology was the large demand for computational resources, nowadays a less significant factor.

The environment effect can be considered a limiting factor by the breeder, hampering selection based exclusively on the phenotypic value. The identification of the best environment to improve the efficiency of a breeding program, mainly as a function of primary components of grain yield, is therefore a challenge for breeders (Hill et al. 1998). Research studies have demonstrated the need to create new selection criteria as a strategy of modifying the conventionally used techniques for winter cereals, focusing on enhanced accuracy and precision in the measurement of genetic differences and environmental effects (Santos and Carvalho 1977, Cruz et al. 1983). Selection applied to quantitative traits in unstable environments deserves more in-depth studies regarding the development of mechanisms enabling the identification of a closer correlation between genotypic and phenotypic values.

In this setting, objectives of this study were i) to estimate the variance components for random factor effects (population) and for the interaction population x sowing method; ii) to perform selection among populations based on the Reml/Blup methodology; iii) to compare the fixed factor effect (sowing method).

MATERIAL AND METHODS

The experiment was installed on an experimental area of the Centro de Genômica e Fitomelhoramento of the Faculdade de Agronomia Eliseu Maciel – Universidade Federal de Pelotas (UFPel), Capão do Leão, state of Rio Grande do Sul, in 2000. The following segregating F_3 oat populations were used in the experiment: UPF 7 x UFRGS 14 (2770), UFRGS 14 x OR 2 (2771), UPF 7 x OR 2 (2772), UFRGS 18x UPF 16 (2773), UFRGS 18 x OR 2 (2774), UPF 16 x OR 2 (2775), UPF 17 x UFRGS 18 (2776), and UFRGS 18 x UPF 14 (2777). These populations were evaluated under three different sowing methods: spaced plants, standard sowing and hill plot. The latter was described as hill method by Frey (1964) and consists of sowing a determined number of seeds per hill. The experiment had a complete randomized block design with two replications. In the hill method, plots were composed of ten hills with 15 seeds per hill, spaced 45 cm apart. In the spaced plant method, plots consisted of ten 2 m long rows spaced 20 cm apart with one plant every 20 cm. In the standard sowing method, plots were composed of a 2 m double row with 65 viable seeds per meter.

After maturation, all panicles from the segregating populations of all eight crosses in both replications and three sowing methods were harvested. The evaluated trait was grain weight panicle^{-1} (PW).

The mixed model or individual Blup (Resende and Fernandes 1999) method was used with a modified estimation of variance components and genetic parameters. Originally, the covariances among parents were estimated and interpreted in terms of their mathematical expectation (i.e. expected values), generating the variance components. Currently, the components of variance can be estimated directly as the variances of random effects of the mixed linear model (Barbosa et al. 2004).

Considering the trait panicle weight in grams (y), the following model was obtained for the phenotypic observations:

$y = \mu + m_i + p_j + pm_{ij} + b_k + e_{ijk}$, where

μ: genotypic or general mean (intercept);

m: effect of the i^{th} method of conducting the populations;

p: effect of the j^{th} segregating population;

pm: effect of the interaction between the i^{th} method of conducting the populations and the j^{th} segregating population;

$b_{k:}$ effect of the k^{th} block;

e_{ijk}: residue effect.

Fixed effects were assumed for the general mean and conduction method, being these effects independent from each other. For the random effects of segregating population (p_j), interaction (pm_{ij}), blocks (b_k), and residue (e_{ijk}) a normal distribution was assumed with mean 0 and variances $^{2}_{p}$, $^{2}_{pm}$, $^{2}_{b}$ and ^{2}, respectively. Additionally, all random effects were assumed to be independent.

The method of restricted maximum likelihood (Reml) in an individual model became the standard method for the estimation of variance components and genetic parameters from

unbalanced data. The superior statistical properties of this method make it preferable to the estimators of least squares and maximum likelihood (Searle et al. 1992). The estimates of components of variance were predicted by the method proposed by Patterson and Thompson (1971).

In its matrix form, the general linear mixed model described by Harville (1977) is shown as follows:

y = Xß + Zí + e

in which

$_{n}$y1 is the vector of observations;

$_{n}X_{p+1}$ is the matrix of fixed effect incidence (known);

$_{p+1}$ß1 is the vector of fixed unknown effects;

$_{n}Z_{q}$ is the matrix incidence of random effects (known);

$_{q}$**í**1 is the vector of random unknown effects;

$_{n}$e1 is the vector of random errors;

where n is the number of observations, p the number of parameters and q the number of random effects.

We assumed that the random effects and errors (residues) have a normal distribution with average zero and were not correlated with the variance and covariance matrices, respectively. **G** and **R** positive defined matrices, by hypothesis, and, therefore, not singular, given by:

Var(**í**) = E(**íí'**) = **G** and Var(**e**) = E(**ee'**) = **R**.

The matrix can be written as:

$$Var \begin{matrix} G & \\ & R \end{matrix}$$

In this sense, we have:

V = Var(y) = Var(Xß) +Var(Zí) + Var(e)=ZVar(í)Z'+R = ZGZ'+R

Assuming, though, that **V** is not singular, and

E(**y**)=E(**Xß + Zí + e)= Xß,**

so

y~N(**Xß; ZGZ'+R).**

RESULTS AND DISCUSSIONS

The effect of fixed factor sowing method indicated significance for F test P > F 0.0001 with 2 (numerator) and 13.6 (denominator) degrees of freedom, so the null hypothesis (H_0) was rejected, as there is at least one contrast of means in this treatment factor that differs from zero.

Table 1 displays the variance component values for panicle weight in oat. The most relevant objective for using the random model, mixed model or type III is the estimation of components directly linked to genetic parameter estimates. By definition, variance components are the variances associated to random effects (Barbin 1995).

Table 1. Variance components of predicted random effects by the method of restricted maximum likelihood (Reml)

Sources of variation	Predicted values	Predicted standard error
Populations	0.02779	0.01795
Populations x Methods	0.0442	0.00630
Blocks	0.0048	0.01220
Error	1.24360	0.01339

Classic and conventional methods used in plant breeding are based on the fact that phenotype is a result of a joint and independent action of genotype and environment, expressed in terms of their variances associated to random factor effects. According to Falconer and MacKay (1996), when quantifying these components, one can reach conclusions on genetic variability by predicting gains to be obtained with the selection of genetically superior or inferior (negative selection) plants and then choose strategies that maximize these gains. Resende and Fernandez (1999) stated that the prediction of genetic values and selection methods depends, essentially, on the estimate of variance components. This means that genetic parameter prediction, i.e. the estimation of the heritability coefficient both in the broad and narrow sense only makes sense for random effects. The method to predict genetic values designated best linear unbiased prediction (Blup) is based on knowledge or a precise estimation of genetic and phenotypic variance components (Henderson 1986).

Analyzing the variance components closely (Table 1) one can see that the large variation between the marginal means of the evaluated segregating population can be attributed to residual variance ($\sigma^2_{\varepsilon} = 1.2436$) which corresponds to over 95% in comparison to genotypic variance ($\sigma^2_{g} = 0.02779$). In practice, the variation in phenotypic expression can be ascribed to the fact that panicle weight is a strictly quantitative trait, therefore determining that the expression of this trait is highly linked to the expression of many genes of small effect on the phenotype.

Most traits of agronomical importance, as for example grain yield and panicle weight, do not present clearly distinct classes, as already observed many years ago by Mendel, i. e., they present continuous variation and can be described as traits of quantitative inheritance.

The first and great postulate of genetics is, generally speaking, that phenotype is a result of individual genotype and environment contributions. In fact, as early as 1909, Johannsen (Mather and Jinks 1984), demonstrated in his experiments with common bean that environmental factors were the major source of variation among traits of quantitative inheritance, leading to the conclusion that phenotype is not a good genotype indicator for this type of trait.

A weighty motive for using mixed linear models is the possibility of predicting random effects in the presence of fixed effects by means of Blup, which is a powerful tool for both

genetics and plant breeding (Dias and Resende 2001). According to Littell et al. (1996), the term prediction refers to random effects and the best linear unbiased prediction can be defined as the result of the regression of effects of a random factor as a function of observation (y) corrected to fixed factor effects.

Table 2 shows the prediction of random effects of eight evaluated segregating F_3 populations. To make these values easier to compare with the obtained adjusted means considering the fixed model, the general mean value of the experiment (2.68622) was added since the random effects are expressed originally in positive and negative values and the sum of mathematical expectation was equal to zero. The highest predicted value found was 2.85573 g for population 2770 and the lowest 2.41093 g for population 2772. The same table shows that the random effects obtained with the mixed model, in comparison to the adjusted means by least squares (LSmeans) are highly correlated. Another fact worth mentioning is that estimates of the predicted genetic value for the segregating populations 2770 (0.16950) and 2772 (-0.2753) stood out as positive and negative values, respectively. For Duchateau and Janssen (1997), Blup represents a contraction of the difference between the marginal and the population means. If the genotypic variance component is much smaller than the environmental variance, the predictor will therefore shrink towards the expected population value (zero). Under such circumstances, very likely there will be no large dispersion among the predicted average genotypic responses.

In the case of narrow genetic variability, the estimates would not be expected to show any variation among the segregating population effects. Therefore, the mixed linear model shows consistency with reality and is seen as a conceptually more complete and informative (Resende and Fernandes 1999) approach. From a practical point of view, it can be concluded that even when the relative genetic variability ($\sigma^2_g/\sigma^2_\varepsilon$) is low, the methodology of mixed linear models can generate selections that are strikingly different when compared to the classic analysis (marginal means and intrablock analysis).

The random effects or genetic values and the genotypic values (adjusted means) for the eight segregating populations

Table 2. Predicted genetic values, genotypic values and adjusted means of eight oat F_3 segregating populations

Population	Random effect		Adjusted means
	Genotypic effects	**Genotypic values** (g)	**Panicle Weight** (g)
2770	0.16950	2.85573	2.7461
2771	-0.14320	2.54303	2.3789
2772	-0.27530	2.41093	2.2222
2773	-0.04727	2.63896	2.4875
2774	0.11860	2.80483	2.6846
2775	-0.02196	2.66427	2.5148
2776	0.07972	2.76595	2.6437
2777	0.11990	2.80613	2.6896

evaluated specifically in each sowing method studied in this work, considering the mixed linear model are displayed in Table 3. For the ease of interpretation of these values, the general mean (model mean, only to have the values expressed in grams) was added to each one of them (Table 3). Regarding the values of random effects, a very distinct performance was observed for the eight evaluated populations, as for example: i) population 2775 evaluated in the standard sowing method (-0.17820), ii) population 2772 evaluated in the hill method (-0.18210), and iii) 2771 (-0.13860) evaluated in the hill method; note that the variance component (P*M) was remarkably lower than both the segregating population variance (0.02779) and the experimental error variance (1.2436). Littell et al. (1996) demonstrated how the values of predictable functions, for example experimental means of random progenies, change when one changes the magnitude of variance components. The interest for information related to the genotypic variance component (σ^2_g) is therefore immediate since it is directly related with the segregating population's genetic potential to produce genetically superior genotypes. Besides, this component is also of general interest, because it is related to the local random variability, i.e., the experimental error variance (σ^2_ε). The conclusion can therefore be drawn that all evaluated segregating populations perform distinctly in the different sowing methods tested.

The analysis of adjusted comparison of means (LSmeans) to the fixed factor (sowing method) was performed using the t test (comparison one to one) for the trait panicle weight, and as a result statistical significance was stated for all evaluated contrasts (Table 4). The spaced plants and standard sowing methods produced the highest (3.0150) and the lowest (2.0970) adjusted means, respectively; in line with the results of Marchioro et al. (2003), who found the highest mean value for panicle weight in grams considering the spaced plants sowing method.

The variance component for populations in each sowing method studied in this experiment is described in Figure 1. Searle et al. (1992) reported that random variable predictors are direct

Table 3. Predicted genetic values and genotypic values (Mg+Blup) for the interaction populations x sowing methods (derivation and panicle weight in grams) of eight oat F_3 segregating populations sown and conducted under the sowing methods: standard sowing, spaced plants and hill plot

Population	Standard Sowing		Spaced Plants		Hill Plot	
	Estimates	PW (g)	Estimates	PW (g)	Estimates	PW (g)
2770	-0.05728	2.62894	0.08719	2.77341	0.05810	2.74432
2771	0.14460	2.83082	-0.08032	2.60590	-0.13860	2.54762
2772	0.09673	2.78295	-0.18210	2.50412	-0.05744	2.62878
2773	-0.02921	2.65701	0.04235	2.72857	-0.03767	2.64855
2774	-0.04671	2.63951	0.12460	2.81082	-0.01644	2.66978
2775	-0.17820	2.50802	0.06047	2.74669	0.10640	2.79262
2776	0.08378	2.77000	-0.04442	2.64180	0.00201	2.68823
2777	-0.01374	2.67248	-0.00778	2.67844	0.08371	2.76993

Table 4. Individually adjusted methods (Least Square Means) for the trait panicle weight (PW) for sowing method factor

Method	PW (g)	IL	SL	P > |t| H0: LSmeans (i)=LSmeans(j)		
				S S	SP	HP
SS	2.0970	1.8848	2.3102	SS	0.0001	0.0001
SP	3.0150	2.8038	3.2252		SP	0.0001
HP	2.5257	2.3129	2.7374			HP

* IL and SL=inferior and superior limit, respectively
SS: Standard sowing; SP: Spaced plant; and HP: Hill Plot

functions of the variance components involved in the model of analysis. Figure 1 shows that this effect once more evidences a narrow variation in the evaluated segregating populations. However, the same figure shows that the interaction population x sowing method effect, if improperly evaluated as adjusted means, would very likely lead to a highly unreliable interpretation. Therefore, random effects predicted through Blups clearly guarantee a higher confidence for these types of estimates, because in this case the biological covariances are taken into account. The use of mixed models can be more appropriate in the case of a high degree of unbalanced data (Robinson 1991). In this study, for example, there were mean effects that consisted of 400 to 1300 observations, making both the conventional statistical analysis and the biological interpretation of the data difficult and unreliable. Nevertheless, Figure 1 shows that each population performed distinctly when evaluated under different sowing methods.

One of the objectives of most plant breeding programs is to estimate the amount of phenotypic variation caused by interaction (Ramalho et al. 2004). With this information on hand, the breeder can direct his work to attenuate its effects. Phenotypic variance can be described as the sum of environmental, genetic and interaction variances among the factors ($\sigma^2_f = \sigma^2_\varepsilon + \sigma^2_g + \sigma^2_{g\varepsilon}$). In this specific case, we have less than 2% of phenotypic variance that can be attributed to interaction effects. There are many sources that frequently contribute to increase experimental error estimates. Ramalho et al. (2000) cite factors such as: i) soil heterogeneity; ii) heterogeneity in the experimental material; iii) plot size and shape; iv) differences in the number of plants in the plot-stand effect; v) treatment type, etc. An estimation of the environmental variance is therefore not sufficient; one must plan the trials in a way that they control the highest possible number of sources of variation.

Besides the complicating factors cited above, one must consider that the number of genes involved in the expression of each trait directly influences the ideal population size required to reveal all possible genotypes, so it is unviable to obtain recombinants for many traits in a single step. For example, if a given trait is determined by eight allele pairs, which is a small number for a quantitative trait, the ideal F_2 generation necessary to reveal all possible genotypes would be 4^8, i.e., 65.536 individuals, a number mostly impossible to obtain in field experiments in view of the limited field area, number of seeds or even financial resources.

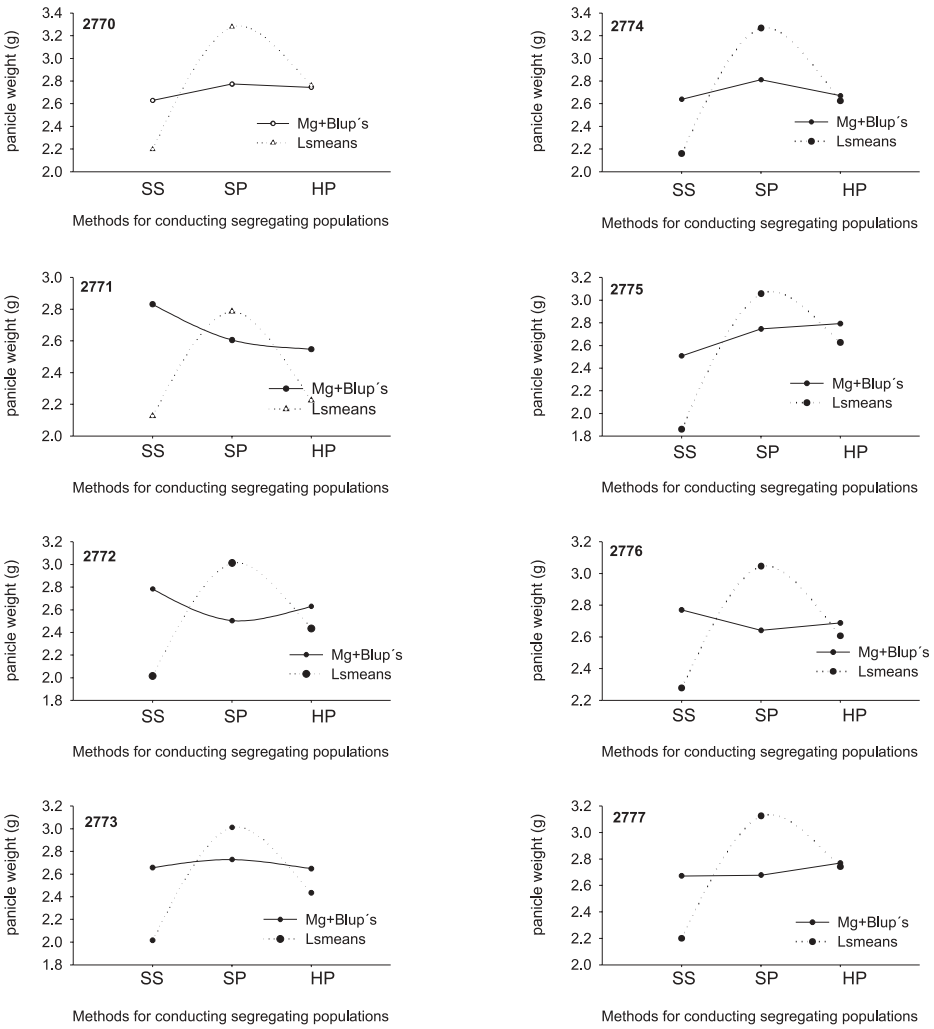

Figure 1. Adjusted means (LSmeans) and general means + random effect (Mg+Blups) of eight oat F_3 segregating populations conducted by three sowing methods (SS: Standard sowing; SP: Spaced plant; and HP: Hill Plot)

Predição do valor genético em populações F_3 de *Avena sativa* L. usando Reml/Blup

RESUMO - *Na área de genética e melhoramento é comum conduzir experimentos constituídos de fatores fixos (método de semeadura) e aleatórios (populações). Sendo assim, a análise estatística mais apropriada deveria ser por meio de modelos lineares mistos. Neste contexto, os objetivos deste trabalho foram: i) estimar os componentes de variância para os efeitos do fator aleatório e interação população x método de semeadura; ii) realizar a seleção entre populações e entre populações dentro de cada método de semeadura; iii) comparar o efeito do fator fixo. Tanto o efeito dos componentes da variância das populações quanto o efeito da interação simples população x métodos de semeadura revelaram uma baixa contribuição para a variância total. Desde modo, pode-se concluir que as populações segregantes apresentam uma estreita variabilidade genotípica. O procedimento Reml/Blup é indicado para a estimação/predição no melhoramento de plantas de aveia, cujos experimentos, em geral, geram dados não balanceados.*

Palavras-chave: lineares mistos**,** predição linear**,** métodos de semeadura.

REFERENCES

Barbin D (1995) **Componentes de variância – Teoria e Aplicações**. São Paulo, Piracicaba, 119p.

Barbosa MHP, Resende MDV, Peternelli LA, Bressiani JA, Silveira, LCI, Silva, FL and Figueiredo ICR (2004) Use of Reml/Blup for the selection of sugarcane families specialized in biomass production. **Crop Breeding and Applied Biotechnology 4**: 218-266.

Bernardo R (1995) Genetic models for predicting maize performance in unbalanced yield trial data. **Crop Science 35**: 141-147.

Bueno Filho JSS and Vencovsky R (2000) Alternativas de análise de ensaios em látice no melhoramento vegetal. **Pesquisa Agropecuária Brasileira 35**: 259-269.

Cruz PJ, Carvalho FIF and Federizzi LC (1983) Efeitos de populações e métodos de seleção aplicados em gerações segregantes de trigo. **Pesquisa Agropecuária Brasileira 18**: 533-541.

Dias LAS and Resende MDV (2001) Estratégias e métodos de seleção. In: Dias LAS (ed.) **Melhoramento genético do cacaueiro**. FUNAPE, Viçosa, p. 217-287.

Duchateau L and Janssen P (1997) An example-based tour in linear mixed models. In: Verbeke G and Molenberghs G (eds.) **Linear mixed models in practice: a SAS-oriented approach**. Springer Series: Lecture notes in Statistics, New York, p. 10-61.

Falconer DS and Mackay TFC (1996) **Introduction to quantitative genetics**. 4th ed., Longman, London, 463p.

Frey KJ (1964) Adaptation reaction of oat strains selected under stress and nonstress environmental conditions. **Crop Science 4**: 55-58.

Harville DA (1977) Maximum likelihood approaches to variances component estimation and to relates problems. **Journal of the American Statistical Association 72**: 320-340.

Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. **Biometrics 31**: 423-448.

Henderson CR (1986) Recent developments in variance and covariance estimation. **Journal of Animal Science 63**: 208-216.

Hill J, Becker HC and Tigersted PMA (1998) **Quantitative and ecological aspects of plant breeding**. Chapman & Hall, London, 288p.

Littell RC, Miliken, GA, Stroup, WW and Wolfinger, RD (1996) **SAS system for mixed models**.: SAS Institute Inc, Cary, 633p.

Marchioro VS, Carvalho FIF, Oliveira AC, Lorencetti C, Benin G, Silva JAG, Cargnin A and Simioni D (2003) Expressão do fenótipo em populações segregantes de aveia conduzidas em diferentes ambientes. **Ciência Rural 33**: 651-656.

Mather K and Jinks JL (1984) **Introdução à genética biométrica**. Revista Brasileira de Genética, Ribeirão Preto, 237p.

Patterson HD and Thompson R (1971) Recovery of inter-block information when block sizes are unequal. **Biometrika 58**: 545-554.

Ramalho MAP and Vencovsky R (1978) Estimação dos componentes da variancia genética em plantas autógamas. **Ciência e Prática 2**: 117-140.

Ramalho MAP, Ferreira DF and Oliveira AC (2000) **Experimentação em genética e melhoramento de plantas**. UFLA, São Paulo, 293p.

Ramalho MAP, Santos JB and Pinto CABP (2004) **Genética na Agropecuária**. 3rd ed., UFLA, São Paulo, 472p.

Resende MDV and Fernandes JSC (1999) Procedimento BLUP individual para delineamentos experimentais aplicados ao melhoramento florestal. **Revista de Matemática e Estatística 17**: 89-107.

Robinson GK (1991) That blup is a good thing: the estimation of random effects. **Statistical Science 6**: 15-51.

Santos FG and Carvalho FIF (1977) Estimativas da seleção para caracteres de importância agronômica em gerações segregantes de trigo (*Tritucum aestivum* L.). **Agronomia Sulriograndense 13**: 219-236.

Scheffé H (1959) **The analysis of variance**. John Wiley, New York, 477p.

Searle S, Casela G and McCulloch CE (1992) **Variance components**. John Willey, New York, 501p.