Designs and model effects definitions in the initial stage of a plant breeding program

Emanuel Fernando Maia de Souza[1], Luiz Alexandre Peternelli[1] and Márcio Henrique Pereira Barbosa[2]

[1]Universidade Federal de Viçosa (UFV), Dep. de Informática, CEP 36570-000 Viçosa, MG. E-mail: emanuelfms@yahoo.com.br, peternelli@dpi.ufv.br [2]UFV, Dep. de Fitotecnia. E-mail: barbosa@ufv.br

Abstract – The objective of this work was to compare the relative efficiency of initial selection and genetic parameter estimation, using augmented blocks design (ABD), augmented blocks twice replicated design (DABD) and group of randomised block design experiments with common treatments (ERBCT), by simulations, considering fixed effect model and mixed model with regular treatment effects as random. For the simulations, eight different conditions (scenarios) were considered. From the 600 simulations in each scenario, the mean percentage selection coincidence, the Pearsons´s correlation estimates between adjusted means for the fixed effects model, and the heritability estimates for the mixed model were evaluated. DABD and ERBCT were very similar in their comparisons and slightly superior to ABD. Considering the initial stages of selection in a plant breeding program, ABD is a good alternative for selecting superior genotypes, although none of the designs had been effective to estimate heritability in all the different scenarios evaluated.

Index terms: augmented design, incomplete block, selection, heritability, sugarcane, stochastic simulation.

Delineamentos e definições de efeitos no modelo em estágios iniciais de melhoramento vegetal

Resumo – O objetivo deste trabalho foi comparar a eficiência relativa da seleção inicial e a estimação de parâmetros genéticos, empregando delineamento em blocos aumentados (DBA), delineamento em blocos aumentados com duas repetições (DBAD) e grupo de experimentos em blocos casualizados com tratamentos comuns (EBCTC), por meio de simulações, considerando o modelo com efeitos fixos e o modelo misto, com tratamentos regulares aleatórios. Trabalhou-se com oito diferentes condições (cenários) com 600 simulações de cada. Foram avaliadas a porcentagem média de coincidência, a estimativa da correlação de Pearson entre as médias ajustadas para o modelo de efeitos fixos e a estimativa da herdabilidade para o modelo misto. O DBAD e o EBCTC foram bastante similares em suas comparações e pouco melhores que o DBA. Considerando os estágios iniciais de seleção em um programa de melhoramento de plantas, o DBA é uma alternativa para seleção dos genótipos superiores, embora nenhum dos delineamentos tenha sido eficaz em estimar herdabilidade em todos os diferentes cenários avaliados.

Termos para indexação: delineamento aumentado, bloco incompleto, seleção, herdabilidade, cana-de-açúcar, simulação estocástica.

Introduction

In order to select genotypes of interest, plant breeders should overcome the obstacles produced at each stage of plant improvement by using dependable techniques to help in the process of choosing the best genotypes. Some difficulties that arise in the initial stages of a sugarcane breeding program are: the need of evaluating a great number of new genotypes that do not show, most of the time, ideal number of stalks for the necessary replications in a complete design; adversities in setting experiments in farms or plants; and the lack of physical and economic resources to carry on the required replications (Peternelli & Barbosa, 2004).

To solve these and other similar problems, Yates (1936), cited by Duarte (2000), proposed the use of incomplete blocks, and Federer (1956) proposed the augmented blocks design (ABD). Pavate (1961), cited by Pimentel-Gomes (1987), proposed the combined analysis of groups of experiments in randomized blocks

with common treatments (ERBCT). These designs retain in common treatments that do not appear in all blocks. Usually, these treatments are referred to as new or regular treatments. In these cases the lack of replication is compensated by mean adjustments (Pimentel-Gomes, 1987).

According to Ramalho et al. (2000), in experiments in randomized blocks with common treatments (ERBCT) the regular treatments are distributed to the various experiments, and within each experiment some common treatments (check) are allocated, according to randomized blocks design. Check treatments are usually well known materials like commercial varieties. Each experiment is a complete block, but together they form an incomplete block design.

Two types of treatments are also considered for ABD: checks and regular. When the blocks are defined, the regular treatments are included only once in just one of the blocks, and the check treatments are included once in all blocks. The repeated treatments (checks) allow error estimation. For ordering, the values of regular treatments are adjusted, as they do not appear in all blocks.

In sugarcane improvement programs, the ABD is frequently used in the initial stages of clone evaluation (Matsuoka, 1999). However, to improve adjusted estimates of the new treatments (genotypes to be evaluated), two replications of the ABD experiment are done, forming duplicated augmented blocks design (DABD), increasing, therefore, the demand for resources and labor force for a same number of clones evaluated (Peternelli & Barbosa, 2004). In this case, analysis of variance is performed with adjusted means of the regular treatments, derived from the analyses of variance of each ABD, considering treatment effects as fixed.

The efficiency of these designs were not investigated when they were proposed hence raising doubts about their effectiveness. With the advances in computer science, it is possible to carry out these studies in the same way as Bearzoti et al. (1997), Duarte (2000) and Santos (2000).

The objective of this work was to compare the relative efficiency for clone selection and genetic parameter estimation of augmented designs and the group of experiments in randomized blocks with common treatments, using simulations, considering a model with fixed effects and another with regular treatment effects as random.

Material and Methods

The simulations were based on the layout of a group of experiments in randomized blocks with common treatments (ERBCT), with three experiments and two replications per experiment, three checks and sixty regular treatments, totaling up to 138 experimental units (Figure 1). Two augmented blocks designs (ABD) were taken from ERBCT: one concerning the first and the other the second replication of all experiments. The duplicated augmented blocks design (DABD) presents the same layout as ERBCT.

For data generation, two coefficients of variation were agreed, 10 and 20% for each error, the residual and among experiments, and coefficient of variation for blocks within the experiment equal to 10%. The heritability values were 0.8 for the check treatments and 0.7 and 0.3 for the regular treatments and 10% selection. The general mean equaled 6.4 units. For each of the eight scenarios, 600 simulations were carried out following the model used for ERBCT. Each group of data generated was analyzed according to each design

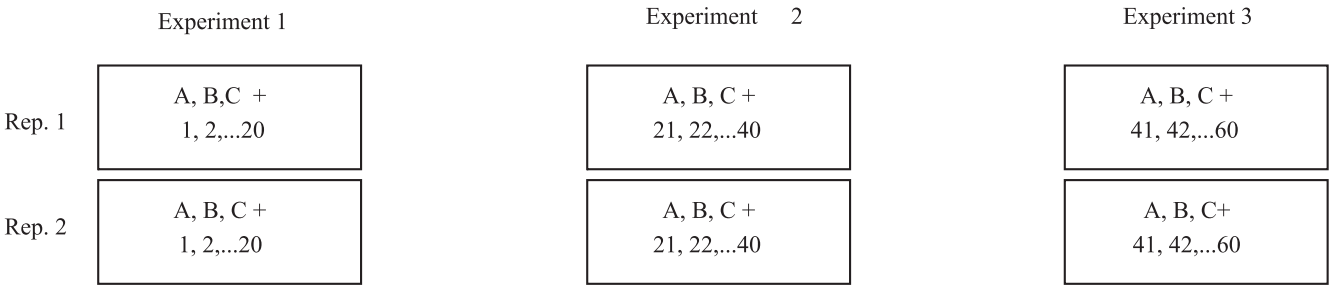

Figure 1. Layout of a group of randomised block design experiments with common treatments (ERBCT), supposing three experiments and two blocks by experiment. Rep. 1: first replication and Rep. 2: second replication. Check treatments: A, B, C. Regular treatments: 1, 2, ...,60.

presented and according to the following statistical models: i) fixed effects model, and ii) mixed model with random regular treatments.

The ERBCT analysis methodology can be found in Pimentel-Gomes (1987) and Ramalho et al. (2000). Different partitions of the total variation can be used, however in the present case, the following fixed effect model was chosen:

$$Y_{ijk} = \mu + E_i + B_{(i)j} + G_k + e_{ijk}, \quad (1)$$

in which Y_{ijk} is the observed value for treatment k, in block j, of experiment i; μ is the general mean; E_i is the effect of experiment i; $B_{(i)j}$ is the effect of block j, within experiment i; G_k is the effect of treatment k; e_{ijk} is the random error associated to the plot that received treatment k, in block j within experiment i.

Due to characteristics of the layout used to perform the simulations, it is important to be aware that when each experiment (exp. i, with i = 1 to 3) from Figure 1 is considered, data from only one block (rep. 1 or rep. 2) would produce, after considering all the experiments, a set of data corresponding to ABD, in which the blocks would actually be one of the accomplishments of each experiment. Each ADB was submitted to the analysis of variance in accordance with model (2) for fixed effects:

$$Y_{jk} = \mu + B_j + G_k + e_{jk}, \quad (2)$$

in which Y_{jk} is the value observed in the experimental plot of block j that received the regular treatment k or the check treatment k' within block j; μ is the general mean; G_k is the effect of treatment k, being k = k' for check treatments, where k' = 1, 2 and 3, and k = 4,..., 63 for regular treatments; B_j is the effect of block, being j = 1, 2 and 3; e_{jk} is the random error associated with the plot of block j, that received check treatment k' or regular treatment k within block j, $e_{jk} \sim NID(0, \sigma^2)$. The means were adjusted as proposed by Federer (1956).

DABD analysis was carried out as follows (Duarte, 2000): the adjusted means of the regular treatments were obtained from each interblock analysis of a particular ABD (corresponding to rep. 1 and rep. 2 of Figure 1), and they were used to compose a new analysis of variance, excluding the check treatments, according to the following model:

$$Y_{ik} = \mu + R_j + G_k + e_{jk}, \quad (3)$$

in which Y_{jk} is the adjusted mean for experimental plot of the regular treatment k (k = 1,..., 60) from replication j (j = 1, 2 and 3); μ is the general mean; R_j is the effect of replication j; G_k is the effect of regular treatment k; and e_{jk} is the random error referring to the experimental plot of the regular treatment k (k = 1,..., 60) from replication j (j = 1, 2 and 3).

The analyses involving random regular treatment effects were performed according to Scott & Milliken (1993), modified by Duarte (2000), in which the effect of treatments is partitioned as $G = T_{k'} + P_{(j)k}$, in which G is the treatment effect; $T_{k'}$ is the fixed effect of check treatment, k' = 1, 2, 3; $P_{(j)k}$ is the random effect of regular treatment k within block j, being k = 4, 5,..., 63; and $P_{(j)k} \sim NID(0, \sigma^2_{T(j)k})$.

Using the adjusted means of regular treatments from each design, ten percent of genotypes with the largest estimates for the variable tested were selected. As the true superior treatments were known (the ones with greater parametric values), it was possible to evaluate the percentages of coincidence among the treatments selected by the designs and the actual superior treatments. The evaluations were performed only with the new treatments, or with the regular treatments.

In the fixed effects model, Pearson's correlation was evaluated (Steel et al., 1997) among the fitted mean values of regular treatments in the three designs, and among the fitted mean values of the regular treatments from each design and the real parametric values of the treatments.

For the random treatment effects model, the broad sense heritability was estimated: $h^2 = \sigma^2_g/(\sigma^2_g+\sigma^2_e)$, in which σ^2_e is the residual variance, and σ^2_g is the genotypic variance estimated by the restricted maximum likelihood (REML) method.

All functions, simulation algorithms and statistical analyses were carried out using R, a free software environment for statistical computing and graphics distributed at R Development Core Team (2004).

Results and Discussion

Each treatment is considered as a different genotype in a plant breeding program. Table 1 shows the percentage of coincidence, in the fixed effects model, and in the mixed model, with regular treatment effects considered as random, among the genotypes selected in the different designs, and among the genotypes selected within the designs and the real best known genotypes. Considering the fixed model, the mean percentage of coincidence among the superior genotypes in ERBCT (V1) varies from 45 to 61%; in ABD (V2) from 36 to 50%; and in DABD, the values equal to ERBCT. By means of a fixed effect model, the fitting expressed by ERBCT and DABD was the same, which is confirmed by variables V4, V5 and V6. The mean percentage of coincidence among the ones selected by ABD and ERBCT or DABD is between 57 and 66%.

Evaluating the design performance within different scenarios, for the mixed model, with treatment effects considered as random, it is observed that they did not present a highly efficient selection in relation to the real superior genotypes (mean percentage of coincidence not exceeding 63% in V1, V2 and V3), however with values, in most cases, a little over the mean coincidences obtained when fixed effects are considered. The mean values of coincidence for ERBCT varied between 48 and 63% (V1); for ABD from 39 to 53% (V2); and for DABD from 47 to 62% (V1). When the evaluation is between designs, the comparison of DABD with ERBCT (V5), representing comparisons between designs with replications of regular treatments, expresses percentage of coincidence around 81% for traits with low heritability and a little over (85%) for traits with higher heritability. When contrasting ABD with ERBCT (V4) and with DABD (V6), the coincidence values obtained are between 55 and 67%, with greater values for the largest heritability and for the V6.

Considering previous evaluations, the adoption of a model with new treatment effects considered to be random, it is possible to improve selection efficiency of the really superior genotypes compared to selection based on mean fittings from the fixed effect model. Duarte (2000) and Santos (2000) observed that the evaluation considering random treatment effects expresses advantages such as the higher accuracy in mean estimation, compared to evaluations considering fixed effects. Selection based on traits with higher heritability can produce better results in any of these models or designs, as expected (Falconer, 1989).

As far as wider residual variation is concerned, on a certain heritability value, the models tend to behave similarly, showing mean coincidences of selection of the genotypes near to each other. Not much variation was observed in the mean percentage of coincidence, within a certain heritability value, among experiments (corresponding to variation among blocks in ABD).

The non-replication of genotypes (use of ABD) caused alteration in coincidence values among the designs. A method to evaluate this alteration, or the ABD efficiency relative to designs with replications (DABD and ERBCT), is through the ratio between the mean percentage of coincidence among the selected by ABD and the real (V2, from Table 1) and the percentage of coincidence among the selected by ERBCT and DABD and the real (V1 or V3, from Table 1). The results are presented in Table 2.

The highest ratio of percentage of coincidence was for ABD/DABD in the model considering random treatment effect, in some cases surpassing 86% (Table 2). Also, the ABD efficiency in relation to the other designs generally ranged from 79 to 87%, even though in this case it uses half area and fewer resources.

If the option is only selecting superior genotypes based on ordering, the use of replications (ERBCT and DABD) does not produce the expected gain in the breeding program in relation to the expenditure of available resources, mainly in the initial stages. This was predictable, since the ABD mean efficiency compared to the other two designs was around 80% (Table 2).

Approaches that consider random treatment effect models derived from different populations may present

Table 1. Mean percent values of coincidence of six variables (V1 to V6) among augmented blocks design (ABD), duplicated augmented block design (DABD), group of randomised block design experiments with common treatments (ERBCT) and the parametric values (real), in the eight scenarios (C1 to C8) with its parameters definitions, in 600 simulations for the fixed model and mixed model with regular treatment effects as random.

Residual CV	Heritability	CV among experiments	Scenarios	Fixed model						Mixed model					
				V1[(1)]	V2	V3	V4	V5	V6	V1	V2	V3	V4	V5	V6
10	0.3	10	C1	47.47	37.57	47.47	57.08	100	57.08	49.92	39.51	47.92	59.18	80.25	55.68
		20	C2	45.78	36.74	45.78	57.71	100	57.71	50.03	39.68	48.00	58.6	81.42	55.40
	0.7	10	C3	59.42	48.94	59.42	64.71	100	64.71	62.11	52.07	60.33	66.86	85.28	64.44
		20	C4	60.67	49.96	60.67	64.46	100	64.46	62.81	52.54	61.22	66.58	85.61	64.00
20	0.3	10	C5	48.67	38.81	48.67	57.60	100	57.60	48.92	39.19	47.39	58.24	80.56	55.57
		20	C6	46.97	37.58	46.97	58.18	100	58.18	49.94	39.75	48.08	58.18	80.64	54.99
	0.7	10	C7	60.67	50.03	60.67	65.75	100	65.75	61.03	51.54	59.50	65.85	85.31	63.69
		20	C8	59.61	49.12	59.61	63.94	100	63.94	61.11	51.50	59.61	66.25	85.81	63.39

[(1)]V1: mean percentage of coincidence between the selected in ERBCT and the real; V2: mean percentage of coincidence between the selected in ABD and the real; V3: mean percentage of coincidence between the selected in DABD and the real; V4: mean percentage of coincidence between the selected in ERBCT and in ABD; V5: mean percentage of coincidence between the selected in ERBCT and in DABD; V6: mean percentage of coincidence between the selected in DABD and in ABD.

changes of treatment ranks for populations of low genotypic variability, when compared to the treatment ranks in approaches that consider fixed effects model (Duarte, 2000). Changes in the classification within a same population are also expected. This may explain the best result presented by models of random treatment effects in the scenarios examined.

Table 3 presents the Pearson's correlation estimates between the values of treatments without environmental effects (actual values) and the fitted means for the fixed effects model in the different designs, as well as the correlation among the fitted means in the different designs.

As the fitted means in the different designs and the actual values of the regular treatments show the same magnitude, Pearson's correlation was used to verify the potential for the recovery of environmental effect in the distinct fitted models given by the different designs. The ABD (V2) fitting for scenarios of lower heritability (C1, C2, C5 and C6) returned relatively low coefficients, near 0.60. The other scenarios offered better correlation, near 0.74. The other designs (V1 and V3) presented, for the same conditions, correlation coefficients close to 0.72 and 0.82, respectively. Another aspect to be observed is that, even with the increase in environmental variation, the mean estimates of correlation were unaffected. The correlation estimates between the designs with replication (DABD and ERBCT) versus the non-replicated one (ABD) were between 0.8 and 0.9, demonstrating that the adjustment given by the mean fitting methods proposed is very close. Also ERBCT and DABD presented the same mean adjustment.

Mean estimates of heritability for each experiment and scenario, as well as the parametric values of heritability used in each simulation are presented in Table 4. It shows that ERBCT was biased toward underestimating the parametric heritability values in the different scenarios. It is also shown that ABD yielded overestimated heritability values for scenarios with lower heritability, and underestimated for the ones with higher heritability. DABD behaved similarly to ABD. It is worth

Table 2. Mean percent values of the ratio between the selected by augmented blocks design (ABD) and the real parametric values of the genotypes (real) and the selected by the group of randomised block design experiments with common treatments (ERBCT) and duplicated augmented block design (DABD) and the real in the eight scenarios (C1 to C8) with its parameters definitions, in 600 simulations.

Residual CV	Heritability	CV among experiments	Scenarios	ABD/ERBCT or ABD/DABD[1]	ABD/ERBCT[2]	ABD/DABD[2]
10	0.3	10	C1	79.14	79.15	82.45
		20	C2	80.25	79.31	82.67
	0.7	10	C3	82.36	83.84	86.31
		20	C4	82.35	83.65	85.82
20	0.3	10	C5	79.74	80.11	82.70
		20	C6	80.01	79.60	82.67
	0.7	10	C7	82.46	84.45	86.62
		20	C8	82.40	84.27	86.39

[1]Models with fixed effects. [2]Model with random treatment effects.

Table 3. Mean coefficients of Pearson's correlation for the fixed effects model in the eight scenarios (C1 to C8) with its parameters definitions, in 600 simulations.

Residual CV	Heritability	CV among experiments	Scenarios	V1[1]	V2	V3	V4	V5
10	0.3	10	C1	0.72	0.60	0.72	0.82	1
		20	C2	0.71	0.59	0.71	0.82	1
	0.7	10	C3	0.84	0.74	0.84	0.88	1
		20	C4	0.84	0.74	0.84	0.88	1
20	0.3	10	C5	0.72	0.59	0.72	0.82	1
		20	C6	0.71	0.59	0.71	0.82	1
	0.7	10	C7	0.84	0.74	0.84	0.88	1
		20	C8	0.84	0.74	0.84	0.88	1

[1]V1: between the fitted means in experiments in randomized blocks with common treatments (ERBCT) and the real genotype values; V2: between the means of the fitted means in augmented block design (ABD) referring to the first (Rep. 1) and second repetition (Rep. 2) and the real genotype values; V3: between the fitted means in duplicated augmented block design (DABD) and the real genotype values; V4: between the means of the fitted means by ABD in Rep. 1 and Rep. 2 and the fitted means in ERBCT; V5: between the means of the fitted means by ABD in Rep. 1 and Rep. 2 and the fitted means in DABD.

Table 4. Mean values of heritability estimated in each design in the eight scenarios (C1 to C8) with its parameters definitions, in 600 simulations[1].

Residual CV	Heritability	CV among experiments	Scenarios	Parametric heritability	ERBCT[1]	ABD1	ABD2	DABD
10	0.3	10	C1	0.3	0.21	0.43	0.41	0.35
		20	C2	0.3	0.21	0.42	0.41	0.35
	0.7	10	C3	0.7	0.44	0.60	0.60	0.55
		20	C4	0.7	0.45	0.60	0.59	0.55
20	0.3	10	C5	0.3	0.27	0.44	0.42	0.34
		20	C6	0.3	0.28	0.45	0.42	0.35
	0.7	10	C7	0.7	0.50	0.60	0.59	0.55
		20	C8	0.7	0.50	0.61	0.60	0.54

[1]ERBCT: experiments in randomized blocks with common treatments; ABD1 and ABD2: augmented block design referring to the first and second repetition; DABD: duplicated augmented block design.

of note that ABD returned mean estimates of heritability nearest to the parametric value (approximately 0.60) when high heritability (0.70) is considered in any experimental condition. When comparing ERBCT to DABD in low heritability conditions (0.30), with a moderate experimental accuracy (CV = 10%), DABD presented the heritability estimates closest to the parametric value (0.30), although overestimated. However, in conditions of low experimental precision (CV = 20%), ERBCT presented heritability estimates closer to the parametric value, when compared to the mean estimates from DABD. ERBCT estimates heritability values more accurately than DABD in experiments with low experimental precision and for traits with low heritability.

Bearzoti et al. (1997), comparing statistical methods to evaluate potato clones, found discrepancies in the heritability estimates of some traits between ABD and lattice designs. Souza et al. (2000) found similar results with beans, reaching the conclusion that ABD would not fit the estimation of genetic and phenotypic parameters due to the low precision of the estimates obtained. However, Santos (2000) argues that the mixed model approach keep advantages compared to fixed effects analysis, mainly when considering random block effects. As for ABD, Duarte (2000) observed that the estimate method and the number of genotypes tested might interfere with estimate quality.

Conclusions

1. ERBCT and DABD present the same mean fitting for the fixed effects model and superior to the mean efficiency of selection by ABD; however, ERBCT estimates heritability values more accurately than DABD in conditions of lower experimental precision and for traits with low heritability.

2. The designs present mean efficiency of selection for random regular treatment effects model slightly higher than the approach with fixed effects.

3. The loss in efficiency of selection by ABD compared to ERBCT and DABD was low, and for characters of high heritability, ABD presents heritability estimates more accurate compared to the other designs, showing that ABD is a good alternative for selection of superior genotypes in the initial stages of a plant breeding program.

Acknowledgements

To Fapemig, for the scholarship and for supporting this study.

References

BEARZOTI, E.; PINTO, C.A.B.P.; OLIVEIRA, M.S. de. Comparação entre métodos estatísticos na avaliação de clones em um programa de melhoramento de batata. **Pesquisa Agropecuária Brasileira**, v.32, p.877-884, 1997.

DUARTE, J.B. **Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal**. 2000. 293p. Tese (Doutorado) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.

FALCONER, D.S. **Introduction to quantitative genetics**. 3rd ed. New York: Wiley, 1989. 438p.

FEDERER, W.T. Augmented (or hoonuiaku) designs. **Hawaian Planters' Record**, v.55, p.191-208, 1956.

MATSUOKA, S.; GARCIA, A.A.F.; ARIZONO, H. Melhoramento da cana-de-açúcar. In: BORÉM, A. (Org.). **Melhoramento de espécies cultivadas**. Viçosa: UFV, 1999. p.205-251.

PETERNELLI, L.A.; BARBOSA, M.H.P. Desafios na condução e planejamento de experimentos de avaliação de clones de cana-de-

açúcar. In: SIMPÓSIO DE ATUALIZAÇÃO EM GENÉTICA E MELHORAMENTO DE PLANTAS, 2004, Lavras. **Desafios e perspectivas da interação biometria e o melhoramento de plantas**: anais. Lavras: Ed. da Ufla, 2004. p.17-27.

PIMENTEL-GOMES, F. **Curso de estatística experimental**. 12.ed. Piracicaba: Esalq/USP, 1987. 467p.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria: Foundation for Statistical Computing, 2004. Available at: <http://www.r-project.org>. Access on: Sept. 2004.

RAMALHO, M.A.P.; FERREIRA, D.F.; OLIVEIRA, A.C. de. **Experimentação em genética e melhoramento de plantas**. Lavras: Ed. da Ufla, 2000. 326p.

SANTOS, A.H. dos. **Teoria de modelos mistos aplicada ao delineamento em blocos aumentados**. 2000. 138p. Dissertação (Mestrado) - Universidade Federal de Lavras, Lavras.

SCOTT, R.A.; MILLIKEN, G.A. A SAS program for analyzing augmented randomized complete-block designs. **Crop Science**, v.33, p.865-867, 1993.

SOUZA, E.A. de; GERALDI, I.O.; RAMALHO, M.A.P. Alternativas experimentais na avaliação de famílias em programas de melhoramento genético do feijoeiro. **Pesquisa Agropecuária Brasileira**, v.35, p.1765-1771, 2000.

STEEL, R.G.D.; TORRIE, J.H.; DICKEY, D.A. **Principles and procedures of statistics**: a biometrical approach. 3rd ed. New York: McGraw-Hill, 1997. 666p.

Received on December 15, 2004 and accepted on February 14, 2006