



# Experimental optimization of the number of blocks by means of algorithms parameterized by confidence interval in popcorn breeding

T.O.M. Paula<sup>1</sup>, C.D. Marinho<sup>2</sup>, A.T. Amaral Júnior<sup>3</sup>, L.A. Peternelli<sup>2</sup> and L.S.A. Gonçalves<sup>3</sup>

<sup>1</sup>Departamento de Fitotecnia, Universidade Federal de Viçosa, Viçosa, MG, Brasil

<sup>2</sup>Departamento de Estatística, Universidade Federal de Viçosa, Viçosa, MG, Brasil

<sup>3</sup>Laboratório de Melhoramento Genético Vegetal, Centro de Ciências e Tecnologias Agropecuárias, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, RJ, Brasil

Corresponding author: C.D. Marinho  
E-mail: [caillet.marinho@yahoo.com.br](mailto:caillet.marinho@yahoo.com.br)

Genet. Mol. Res. 12 (2): 2109-2119 (2013)

Received August 14, 2012

Accepted May 2, 2013

Published June 27, 2013

DOI <http://dx.doi.org/10.4238/2013.June.27.2>

**ABSTRACT.** The objective of this study was to determine the optimal number of repetitions to be used in competition trials of popcorn traits related to production and quality, including grain yield and expansion capacity. The experiments were conducted in 3 environments representative of the north and northwest regions of the State of Rio de Janeiro with 10 Brazilian genotypes of popcorn, consisting by 4 commercial hybrids (IAC 112, IAC 125, Zélia, and Jade), 4 improved varieties (BRS Ângela, UFVM-2 Barão de Viçosa, Beija-flor, and Viçosa) and 2 experimental populations (UNB2U-C3 and UNB2U-C4). The experimental design utilized was a randomized complete block design with 7 repetitions. The Bootstrap method was employed to obtain samples of all of the possible combinations within

the 7 blocks. Subsequently, the confidence intervals of the parameters of interest were calculated for all simulated data sets. The optimal number of repetition for all of the traits was considered when all of the estimates of the parameters in question were encountered within the confidence interval. The estimates of the number of repetitions varied according to the parameter estimated, variable evaluated, and environment cultivated, ranging from 2 to 7. It is believed that only the expansion capacity traits in the Colégio Agrícola environment (for residual variance and coefficient of variation), and number of ears per plot, in the Itaocara environment (for coefficient of variation) needed 7 repetitions to fall within the confidence interval. Thus, for the 3 studies conducted, we can conclude that 6 repetitions are optimal for obtaining high experimental precision.

**Key words:** Experimental design; Optimal number of repetitions; *Zea mays*

## INTRODUCTION

In the improvement of plants, special attention should be given to experimental precision, since imprecision commonly causes poor discrimination between treatments and results in erroneous conclusions in relation to the genotypes that should be selected (Cargnelutti Filho and Storck, 2007; Storck et al., 2009).

Among the factors that affect experimental precision, the appropriate use of the number of repetitions assumes relevant function. It is well known that, to some extent, the greater the number of repetitions, the greater the precision that will be obtained and the better the experimental estimates that can be achieved (Steel and Torrie, 1980). However, the utilization of a large number of repetitions may restrict the quantity of materials evaluated, because it involves increase in the size of the experiment and, consequently, more resources needed to establish and conduct these experiments.

There are several other factors that affect experimental precision, among them: the heterogeneity of the soil, the genetic material, number of environments utilized, sample size, and the management of the experiment (Martin et al., 2005; Storck et al., 2006; Brum et al., 2008; Cargnelutti Filho et al., 2008).

Various studies about experimental precision are reported in the literature in which the cultures utilized were: maize (Martin et al., 2005); sorghum (Brum et al., 2008); potato (Storck et al., 2011); soybean and bean, castor bean, crambe, turnip, rice (Cargnelutti Filho et al., 2009, 2010, 2011a,b, 2012, respectively); peach (Toebe et al., 2012).

However, Cattapatti et al. (2008) reported the only Brazilian study about experimental precision with popcorn, concluding that for pre-harvest traits, one can use 4 repetitions and 5 plants per plot. Therefore, lacking research on experimental information about the agricultural practices for popcorn, the same management strategies recommended for maize are being adopted (Oliveira et al., 2003; Cattapatti et al., 2008). To this end, the objective of this study was to determine the optimal number of repetitions to be utilized in competition trials of popcorn for traits related to production and quality.

## MATERIAL AND METHODS

The experiments were conducted in the agricultural year of 2007/2008, at the Experimental Station of PESAGRO-RIO (in the cities of Campos dos Goytacazes and Itaocara, RJ) and at Colégio Estadual Agrícola Antônio Sarlo (Campos dos Goytacazes, RJ), thus making 3 environments.

Each trial consisted of 10 genotypes (BRS Ângela, UFVM-2 Barão de Viçosa, IAC 112, IAC 125, Zélia, Jade, Beija-Flor, Viçosa, UNB2U-C3, and UNB2U-C4) and the design utilized was a randomized complete block design with 7 repetitions. The plot was composed of a grid of 5.00 m, with rows spaced 0.90 m apart with 25 plants per plot, planted equidistant at 0.02 m.

The following traits were evaluated: average height of the plant, in cm; average height of the insertion of the first ear, in cm; number of ears per plot; grain yield, in kg/ha (GY); flowering time in days; weight of 100 kernels, in g; expansion capacity (EC), in mL/g.

Samples to be compared were acquired by obtaining all possible combinations among the 7 blocks evaluated, using the bootstrap resampling technique (Davison and Hinkley, 1997). Therefore, random samples were removed from the set of 7 blocks so that the sampled blocks would never be repeated within the same sampling. Five sample sizes were tested with N equal to 2, 3, 4, 5, and 6 blocks; resampling with replacement was performed, thus generating a dataset for each value of N in every variable analyzed, totaling 120 sets for each environment.

One-way analysis of variance was performed for each dataset and from these analyses the parameter estimates of interest were obtained to verify the effect of the different number of repetitions.

Analysis of variance was performed considering the following statistical model:  $Y_{ij} = \mu + R_j + G_i + \xi_{ij}$ , where  $\mu$  is the overall average,  $R_j$  is the  $j^{\text{th}}$  repetition effect,  $G_i$  is the fixed effect of  $i^{\text{th}}$  genotype, and  $\xi_{ij}$  is the experimental error (Cruz et al., 2004).

The parameters estimated were obtained according to the following expressions:  $\hat{\Phi}_g = \frac{MSG - MSE}{r}$  (genotypic variability);  $\hat{\sigma}_f^2 = \frac{MSG}{r}$  (phenotypic variance);  $\hat{\sigma}_r^2 = \frac{MSE}{r}$  (environmental or residual variance);  $\hat{H}^2 = \frac{\hat{\Phi}_g}{\hat{\sigma}_f^2}$  (genotypic determination coefficient), where  $\hat{\Phi}_g = \frac{\sum G_i^2}{g-1}$ ; and  $\hat{C}V_c(\%) = (100 \cdot \sqrt{MSE} / \bar{x})$  (experimental coefficient of variation), where  $MSG$  is the mean square of genotype,  $MSE$  is the mean square of residual, and  $r$  is the number of repetitions.

The confidence intervals (CIs) for the values of the parameters were estimated based on the original dataset (7 blocks). We used the methodology recommended by Knapp et al. (1985) for the calculation of the CI for the genotypic coefficient of determination, which was obtained using the following expression:

$$CI_{1-\alpha} = \left\{ 1 - \left[ \left( \frac{M_1}{M_2} \right) F_{1-\alpha/2; g/2, g/1} \right]^{-1}; 1 - \left[ \left( \frac{M_1}{M_2} \right) F_{\alpha/2; g/2, g/1} \right]^{-1} \right\}$$

where,  $M_1$  and  $M_2$  are the values estimated for MSG and MSE, respectively.

To obtain the CI for the experimental coefficients of variation, we used the algorithm proposed by Vangen (1996):

$$CI_{1-\alpha/2} = \left[ \frac{CV}{\sqrt{\left(\frac{U_2+2}{v+1}-1\right)CV^2 + \frac{U_2}{v}}}; \frac{CV}{\sqrt{\left(\frac{U_1+2}{v+1}-1\right)CV^2 + \frac{U_1}{v}}} \right],$$

where,  $U_1$  and  $U_2$  are the upper quantiles  $1-\alpha/2$  and  $\alpha/2$  of the chi-squared distribution with  $v = n-1$  degrees of freedom.

For the genotypic variability, phenotypic variance, and residual variance, the CIs were obtained by the methodology described by Barbin (1993), in which the lower and upper limits of the confidence interval are given by  $IL = \frac{DF \times \hat{\phi}_g}{X_{\alpha/2}^2}$  and  $SL = \frac{DF \times \hat{\phi}_g}{X_{1-\alpha/2}^2}$ , where  $DF$  is the degrees of freedom associated with the variance components estimated, according to Satterthwaite (Barbin, 1993), and  $X_{\alpha/2}^2$  and  $X_{1-\alpha/2}^2$  are the upper quantiles of the chi-squared distribution  $\alpha/2$  and  $1-\alpha/2$ , with  $\alpha = 5\%$  of probability.

The optimal number of repetition for all of the traits was considered when all the parameter estimates in question were found within the CI. All statistical analyses were performed using the R statistical package (R Development Core Team, 2006).

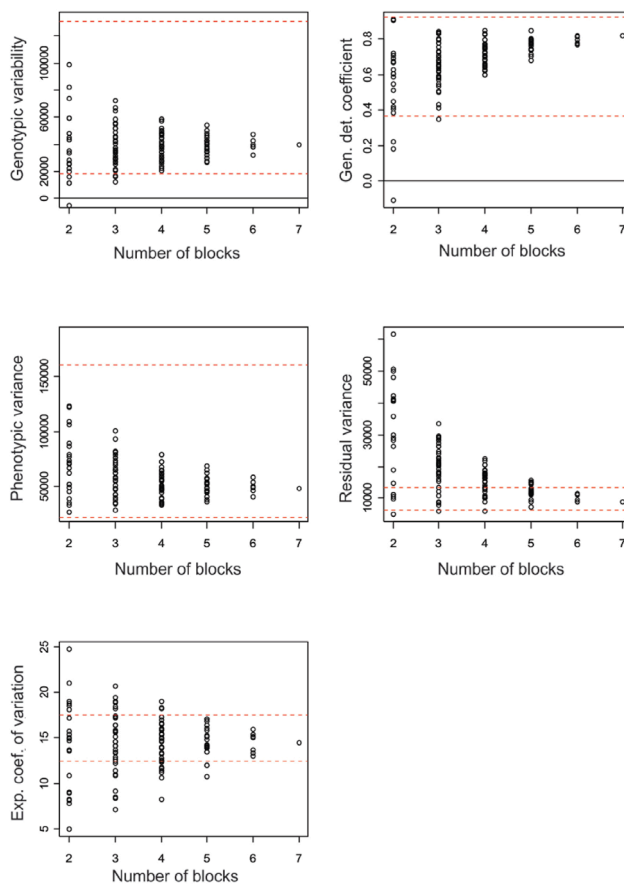
## RESULTS AND DISCUSSION

The dispersions of the parameter estimates for the variables GY and EC in the 3 environments evaluated are presented in Figures 1-6. These traits were highlighted as they were considered the most important among the traits evaluated, and the information about the other traits is compiled in Table 1.

The dashed lines in the figures indicate the upper and lower limits of the CI drawn for the values of the estimated parameters in the set of 7 blocks. The general trend was as expected, that is, the variation of the parameter estimates increased as the number of blocks decreased (Figures 1-6), thus indicating the reduction of the experimental efficiency with a decrease of repetitions (blocks).

In the Colégio Agrícola environment, it was noted that, for genotypic variability and genotypic determination coefficient, 4 blocks were ideal for GY (Figure 1). On the other hand, for the phenotypic variance with 2 blocks the estimates were all included within the CI, this being the number of repetitions indicated (Figure 1).

Considering the distribution of the estimates and the established intervals for the residual variance and coefficient of experimental variation (CVe), we conclude that 6 was the optimal number of blocks and that there was thus a correct evaluation of the genotypes.



**Figure 1.** Dispersion of the estimate of the genotypic variability, genotypic determination coefficient, phenotypic variance, residual variance, and experimental coefficient of variation for the grain yield in the Colégio Agrícola environment in Campos dos Goytacazes.

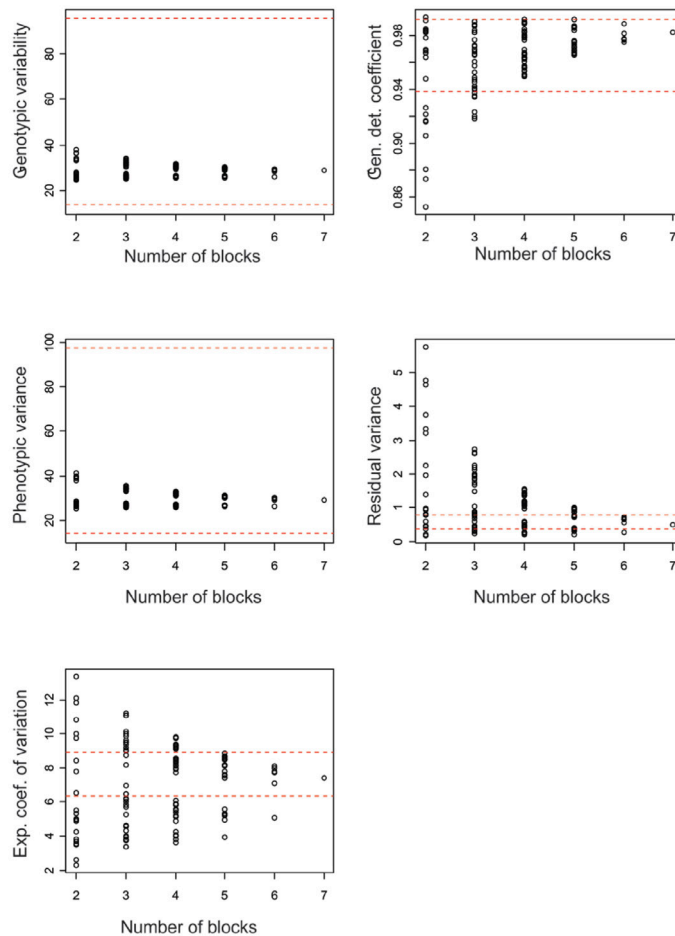
**Table 1.** Optimal number of repetitions (blocks) for the estimation of parameters, obtained via Bootstrap and confidence intervals, in the height of plant (HP), height of ear (HE), number of ears per plot (NE), grain yield (GY), flowering time in days (FT), weight of 100 kernels (W100), expansion capacity (EC) traits in 10 genotypes of popcorn maize, in the environments of ColégioAgrícola (Campos dos Goytacazes) and PESAGRO-RIO (Campos dos Goytacazes and Itaocara).

Parameters	Colégio Agrícola							Itaocara							PESAGRO						
	HP	HE	NE	GY	FT	W100	EC	HP	HE	NE	GY	FT	W100	EC	HP	HE	NE	GY	FT	W100	EC
$\hat{\phi}_g$	4	5	6	4	5	3	2	4	6	6	4	3	2	2	2	2	3	4	2	2	2
$\hat{H}^2$	4	4	4	4	4	4	6	4	4	4	4	5	4	3	4	4	4	4	4	4	4
$\hat{\sigma}_f^2$	3	5	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
$\hat{\sigma}_r^2$	6	6	6	6	6	6	7	6	6	6	6	6	6	6	6	6	6	6	6	6	6
$C\hat{V}_e$	6	6	5	6	6	6	7	5	5	7	6	5	5	6	6	6	5	6	5	6	6

$\hat{\phi}_g$  = genotypic variability;  $\hat{H}^2$  = genotypic determination coefficient;  $\hat{\sigma}_f^2$  = phenotypic variance;  $\hat{\sigma}_r^2$  = environmental or residual variance;  $C\hat{V}_e$  = experimental coefficient of variation.

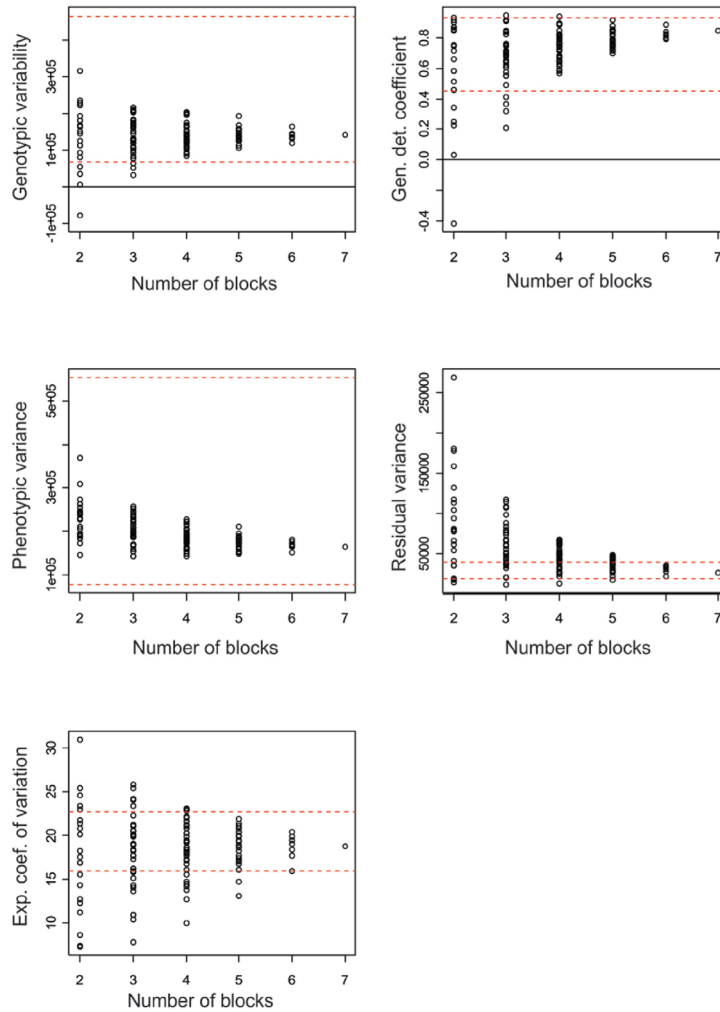
Catapatti et al. (2008) concluded that for pre-harvest traits and productivity, one can sample 5 plants and use 4 repetitions, and emphasized that there was a greater number of significant differences between averages of genotypes, in the numbers of repetitions more than 4.

For the EC trait, even in the Colégio Agrícola environment, the optimal number of repetitions for genotypic variation and phenotypic variance was 2, while for the coefficient of variation, the optimal number of repetitions was 6, and for residual variance and CVe, the optimal number of repetitions was 7 (Figure 2).



**Figure 2.** Dispersion of the estimate of the genotypic variability, genotypic determination coefficient, phenotypic variance, residual variance, and experimental coefficient of variation for the expansion capacity in the Colégio Agrícola environment in Campos dos Goytacazes.

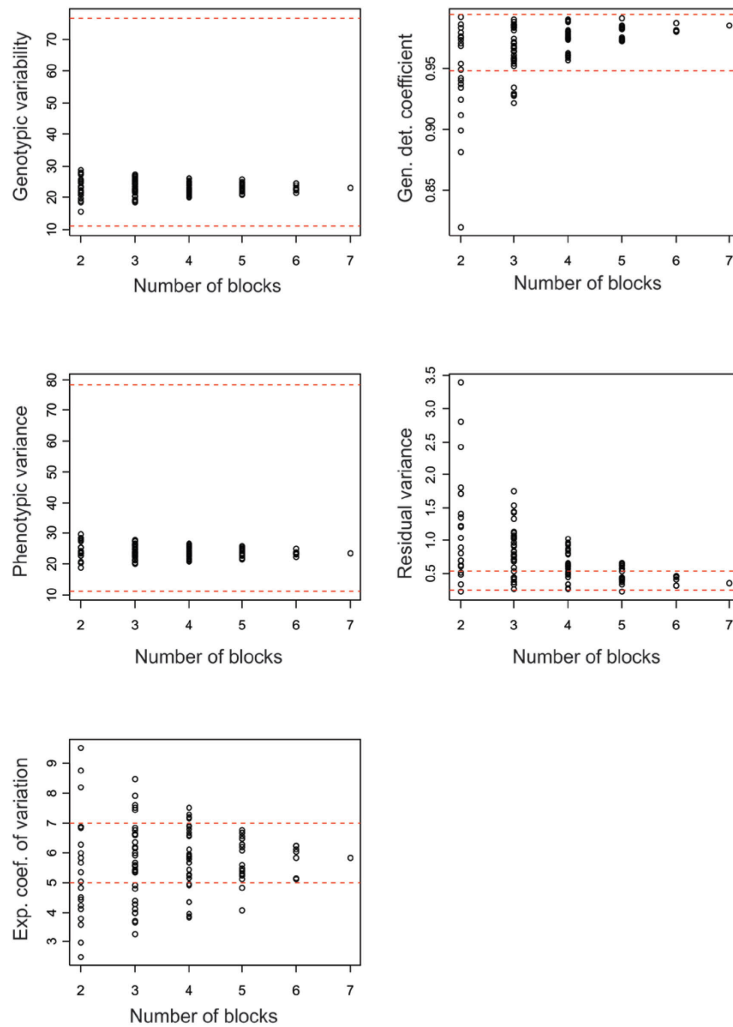
The graphs of the dispersion of the estimates for GY at PESAGRO-RIO (Itaocara) reveal that 6 blocks were sufficient to precisely estimate all of the parameters proposed for this trait (Figure 3). However, separately considering the phenotypic variance, genotypic variability, and genotypic determination coefficient, 2, 5, and 4 repetitions were sufficient, respectively.



**Figure 3.** Dispersion of the estimate of the genotypic variability, genotypic determination coefficient, phenotypic variance, residual variance, and experimental coefficient of variation for the grain yield in the PESAGRO-RIO in Itaocara environment.

Resende and Duarte (2007) found that the number of repetitions usually employed (2 to 4) for production traits, generally with low genotypic determination coefficients, is not enough, and recommended at least 6 repetitions in the evaluation of these traits, to achieve higher selective accuracy.

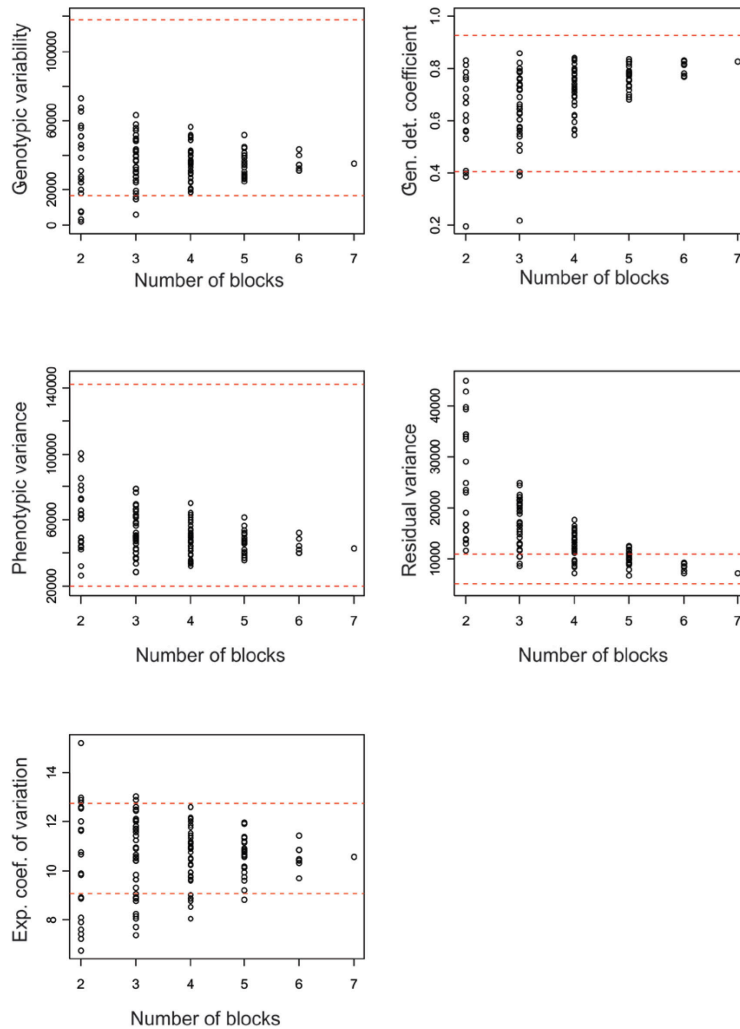
In analyzing the EC trait at PESAGRO-RIO (Itaocara), 2 repetitions were found to be optimal for genotypic variation and phenotypic variance. For the genotypic determination coefficient, the optimal number was 4 repetitions, while for residual variance and C<sub>V</sub>e, it was 6 (Figure 4).



**Figure 4.** Dispersion of the estimate of the genotypic variability, genotypic determination coefficient, phenotypic variance, residual variance, and experimental coefficient of variation for the expansion capacity in the PESAGRO-RIO in Itaocara environment.

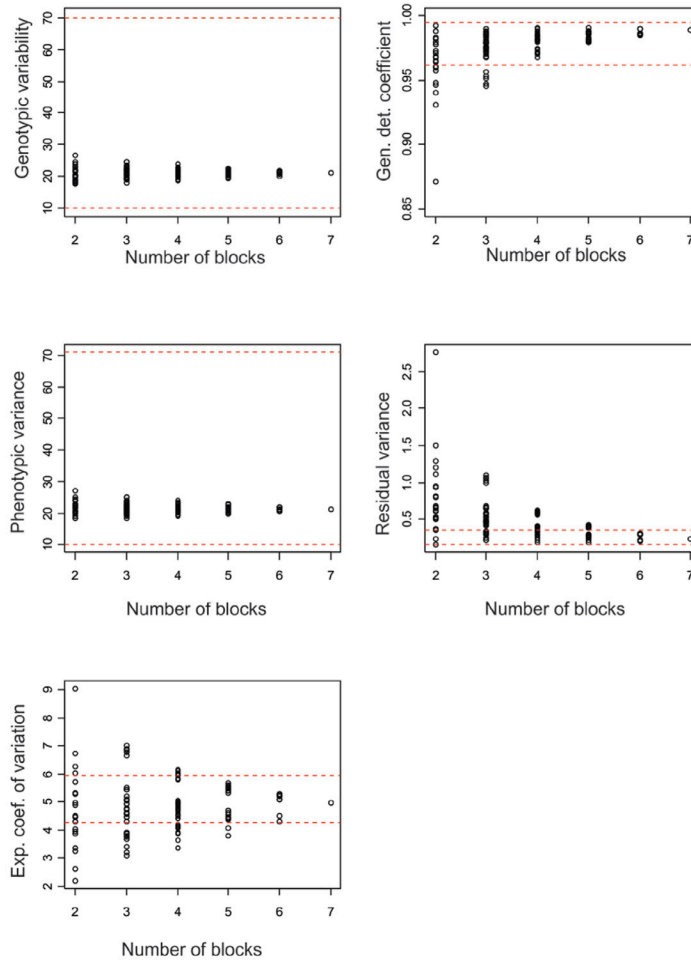
For the distribution of the estimates of the parameters for the variable GY in the environment PESAGRO-RIO (Campos dos Goytacazes), 4 was the optimal number of blocks for inferences about genotypic variability and the genotypic determination coefficient. Two repetitions would be optimal for phenotypic variance, and 6 repetitions for residual variance CVe (Figure 5). For EC, the optimal number of repetitions was 2 (genotypic variability and phenotypic variance), 4 (genotypic determination coefficient) and 6 (residual variance and CVe) (Figure 6).





**Figura 5.** Dispersion of the estimate of the genotypic variability, genotypic determination coefficient, phenotypic variance, residual variance, and experimental coefficient of variation for the grain yield in the PESAGRO-RIO in Campos dos Goytacazes environment.

Note that for the 3 environments, the genotypic determination coefficient for EC showed higher estimates. According to Ramalho et al. (2005), the increase in the number of repetitions for traits with high heritability does not offer much benefit to the breeder, but when it comes to heritability lower than 50%, the use of a greater number of repetitions advantageously increases the expected gain with selection, and moreover the experimental precision of course. Therefore, considering only the genotypic determination coefficient, one can consider reducing the optimal number of repetitions without great losses in the inferences.



**Figura 6.** Dispersion of the estimate of the genotypic variability, genotypic determination coefficient, phenotypic variance, residual variance, and experimental coefficient of variation for the expansion capacity in the PESAGRO-RIO in Campos dos Goytacazes.

The optimal number of repetitions (blocks) for all traits evaluated in this study, including GY and EC, are compiled in Table 1. Note that for the genotypes evaluated, the optimal number of repetition varies according to the parameter in question, the environment, and the variable being evaluated as well.

It is believed that only the EC traits in the Colégio Agrícola environment (for residual variance and C<sub>Ve</sub>), and number of ears per plot, in the Itaocara environment (for C<sub>Ve</sub>) needed 7 repetitions to fall within the CI. Thus, considering the parameters estimated and the traits evaluated, we can conclude that 6 repetitions are optimal for obtaining high experimental precision in the 3 studies conducted.

## REFERENCES

- Barbin D (1993). Componentes de Variância: Teoria e Aplicações. Fundação de Estudos Agrários Luiz de Queiroz, Piracicaba.
- Brum B, Lopes SJ, Storck L, Santos VJ, et al. (2008). Tamanho ótimo de parcela para ensaios com sorgo granífero em duas épocas de semeadura. *Ciênc. Rural* 38: 315-320.
- Cargnelutti Filho A and Storck L (2007). Estatísticas de avaliação da precisão experimental em ensaios de cultivares de milho. *Pesq. Agropec. Bras.* 42: 17-24.
- Cargnelutti Filho A, Ribeiro ND, Reis RCP, Souza JR, et al. (2008). Comparação de métodos de agrupamento para o estudo da divergência genética em cultivares de feijão. *Ciênc. Rural* 38: 2138-2145.
- Cargnelutti Filho A, Storck L and Ribeiro ND (2009). Medidas da precisão experimental em ensaios com genótipos de feijão e de soja. *Pesq. Agropec. Bras.* 44: 1225-1231.
- Cargnelutti Filho A, Lopes SJ, Brum B, Silveira TR, et al. (2010). Tamanho de amostra de caracteres em híbridos de mamoneira. *Ciênc. Rural* 40: 250-257.
- Cargnelutti Filho A, Lopes SJ, Toebe M, Silveira TR, et al. (2011a). Tamanho de amostra para estimação do coeficiente de correlação de Pearson entre caracteres de *Crambe abyssinica*. *Rev. Ciênc. Agronom.* 42: 149-158.
- Cargnelutti Filho A, Toebe M, Burin C, Fick AL, et al. (2011b). Tamanhos de parcela e de ensaio de uniformidade em nabo forrageiro. *Ciênc. Rural* 41: 1517-1525.
- Cargnelutti Filho A, Marchesan E, Silva LS and Toebe M (2012). Medidas de precisão experimental e número de repetições em ensaios de genótipos de arroz irrigado. *Pesq. Agropec. Bras.* 47: 336-343.
- Catapatti TR, Gonçalves MC, Silva Neto MR and Sobroza R (2008). Tamanho de amostra e número de repetições para avaliação de caracteres agronômicos em milho pipoca. *Ciênc. Agrotecnol.* 32: 855-862.
- Cruz CD, Regazzi AJ and Carneiro PCS (2004). Modelos Biométricos Aplicados ao Melhoramento Genético. Editora UFV, Viçosa.
- Davison AC and Hinkley DV (1997). Bootstrap Methods and their Application. Cambridge, New York.
- Knapp SJ, Stroup WW and Ross WM (1985). Exact confidence intervals for heritability on a progeny mean basis. *Crop Sci.* 25: 192-194.
- Martin TN, Storck L, Dal'Col Lúcio A, Carvalho MP, et al. (2005). Bases genéticas de milho e alterações no plano experimental. *Pesq. Agropec. Bras.* 40: 35-40.
- Oliveira RH, Oliveira FA, Santos AC and Jacome AG (2003). Rendimento e componentes de produção de milho pipoca em função de resíduos de adubação e densidade populacional. *Cult. Agrônôm.* 12: 89-102.
- R Development Core Team (2006). R: A Language and Environment for Statistical Computing. Available at [<http://www.r-project.org/>]. Accessed January 20, 2010.
- Ramalho MAP, Ferreira DF and Oliveira AC (2005). Experimentação em Genética e Melhoramento de Plantas. Editora UFLA, Lavras.
- Resende MDV and Duarte JB (2007). Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesq. Agropec. Trop.* 37: 182-194.
- Steel RGD and Torrie JH (1980). Principles and Procedures of Statistics: A Biometrical Approach. McGraw-Hill, New York.
- Storck L, Martin TN, Lúcio AD, Lopes SJ, et al. (2006). Tamanho ótimo de parcela em experimentos com milho relacionado a metodologias. *Rev. Bras. Milho Sorgo* 5: 48-57.
- Storck L, Cargnelutti Filho A, Lopes SJ, Toebe M, et al. (2009). Duração do subperíodo semeadura-florescimento, crescimento e produtividade de grãos de milho em condições climáticas contrastantes. *Rev. Bras. Milho Sorgo* 8: 27-39.
- Storck L, Lopes SJ, Lúcio AD and Cargnelutti Filho A (2011). Optimum plot size and number of replications related to selective precision. *Ciênc. Rural* 41: 390-396.
- Toebe M, Both V, Brackmann A, Cargnelutti Filho A, et al. (2012). Tamanho de amostra para a estimação da média de caracteres de pêssego na colheita e após o armazenamento refrigerado. *Ciênc. Rural* 42: 209-212.
- Vangen MG (1996). Confidence interval for a normal coefficient of variation. *Am. Stat.* 50: 21-26.