

SAMPLE SIZE FOR ESTIMATING THE WITHIN PLOT VARIANCE IN EXPERIMENTS DESIGNED TO EVALUATE COMMON BEAN PROGENIES

Elaine Aparecida de Souza, Daniel Furtado Ferreira and
Magno Antonio Patto Ramalho

ABSTRACT

Ninety-three segregant progenies, of the F₃, F₄ and F₅ generations, from the cross between common bean cultivars Jalo and Small White, were assessed during the dry season of 1991. The purpose was to calculate the within plot variability and to determine the minimum plot sample size necessary to quantify this variation. For these experiments five plants from each plot were harvested individually and the mean and variance of yield, number of pods and number of seeds per plant were calculated. Different sample sizes were simulated, varying from 5% to 50% of the number of plots. Ten samples of the same size were used for each simulation. By comparing the estimates obtained per sample with those of the whole population, it was shown that the samples were representative of the population for all characteristics and generations even when the smallest size, 5% of the plots, was used.

INTRODUCTION

Evaluation is the most important step in the breeding program of any crop, requiring great care and resources. Also, data collection is time consuming due to the large amount of material normally scored.

Evaluations can be started from the F₃ or F₄ generation, when there is still a lot of variation within families (Ramalho and Vencovsky, 1978; Fehr, 1987). To quantify

this variation, it is necessary to obtain information at the level of individuals. As the number of families is large, usually more than 100, with few replications and a relatively large number of individuals within a plot, the collection of this information at the level of the individual is nearly always impossible, due to the need to handle thousands of individual plants.

To minimize this problem, an alternative is to collect only one sample representative of the variation within all the families. In some crops, such as maize, this practice has been adopted on some occasions (Pacheco, 1987). However, there are no reports, not even for maize, of how the sample size was determined. In the case of the common bean, no reference to such a procedure was found. Thus this work was carried out to verify the possibility of using one sample from the plots to quantify the variation within the progenies, and the minimum sample size which would be representative of this variation.

MATERIAL AND METHODS

The varieties used were "Jalo", which has large yellow-colored seeds and "Small White" with small white seeds. They were crossed in 1988, in Lavras, using the methodology presented by Vieira (1967). The F₁ and F₂ generations were obtained and 93 F₂ plants were selected. The families were managed by within family bulk breeding from the F₂ to the F₅ generation.

The F₃, F₄ and F₅ generations were simultaneously assessed in 1991. For the F₃ families, a separate experiment was carried out in a 10x10 lattice design with two replications. In this group parents and five controls were included with the 93 F₃ progenies. The plot was made up of a one meter single row, 10 seeds sown per linear meter. The F₄ and F₅ progenies were assessed in a different experiment, using a 14x14 lattice, with three replications. The plot was made up of a two meter row, with 0.5 m spacing and 15 seeds sown per linear meter.

Five plants were randomly harvested from each plot. The number of pods and seeds and the grain weight of each plant were recorded. The mean and variance of each plot were calculated and from these data the within plot phenotypic variance ($\hat{\sigma}_d^2$) was estimated. Different sample sizes, varying from 5% to 50% of the number of plots, were simulated. For each sample size in this simulation, ten different samples with replacement were obtained. The average within plot variance ($\hat{\sigma}_d^2$) was calculated for each sample. The smallest and greatest values were identified from these calculations and the average variance was estimated.

RESULTS AND DISCUSSION

The within plot variance estimates varied between characteristics and generations. For the characters, this variation can be explained by distinct genetic segregation and by the action of the environment.

The F₄ and F₅ generations were sown in a single experiment, with the same sowing density. The F₃ generation, in spite of being obtained in the same agricultural year, was sown with a lower density. This contributed to a greater productivity of the F₃ families and reflected in the calculation of the within plot variance, which in the case of the yield per plant, was about three times greater than that obtained in the F₄ and F₅ generations. The difference in the number of plots sampled in the F₄ and F₅ generation was due to the loss of some plots. For the F₃ generation the number was smaller because the assessment included only two replications and because there was also a missing plot.

Taking the yield of beans per plant of the F₃ generation as a reference (Table I), it can be seen that the average within plot variance of the 10 samples for each sample size did not differ. They also did not differ from the average variance of all the plots. Apparently even a sample as small as 5% of the plots can supply a good estimate of the within plot variance.

The range of the estimates of $\hat{\sigma}_d^2$ between samples of the same size was much greater when smaller proportions of the plots were sampled. For the samples of 5% to 10% of the total plots the amplitudes were larger. Nevertheless, for samples of 20% or more of the selected plots the amplitudes of these variations were small.

When the number of pods per plant of the same F₃ generation was considered, similar results were observed. However in the sample sizes of 5% and 10% the within plot mean variance was slightly greater than that of the whole population (Figure 1). The result for the number of the seeds per plant was similar to that for grain yield per plant (Figure 2).

This sampling effect was similar in the different generations. The same procedure was used for the F₄ and F₅ generations (Table I), and although the estimates of $\hat{\sigma}_d^2$ were lower than those of the F₃ generation, for the reasons already mentioned, the sampling effect was similar.

In order to determine if the average variances, estimated for each sample size ($\hat{\sigma}_d^2$), differed from the calculation of $\hat{\sigma}_d^2$, when all the plots are considered, the relation $\hat{\sigma}_d^2 / \hat{\sigma}_d^2$ (Table I) was calculated. The values of this ratio, in all cases, proved that the average within plots variances were practically the same, since they were close to 1.0 and showed a very small range of variation.

From these data, it is evident that if the plant breeder wishes to get information on within plot variation, for the character grain yield, he can get the information at the

Table 1 - Estimate of the phenotypic within plot variance ($\overline{\sigma_d^2}$) for grain yield per plot with different sample sizes, and the range of variation of these estimates in the F₃, F₄ and F₅ generations of the common bean.

Sample size (%)	F ₃			F ₄			F ₅					
	$\overline{\sigma_d^2}$	Low	Hig	$\overline{\sigma_d^2}/\sigma_d^2$	Low	Hig	$\overline{\sigma_d^2}/\sigma_d^2$	Low	Hig	$\overline{\sigma_d^2}/\sigma_d^2$		
5	33.2	16.9	45.0	1.02	11.3	7.9	14.6	1.02	9.7	5.6	13.9	0.83
10	33.6	20.3	44.1	1.04	10.7	6.6	15.5	0.97	12.3	6.5	22.3	1.05
15	33.4	26.2	41.4	1.03	11.3	7.5	15.4	1.02	10.4	7.6	17.5	0.89
20	33.9	28.7	44.6	1.05	11.8	8.8	17.0	1.07	11.9	8.1	16.6	1.02
25	32.6	26.6	42.2	1.01	10.9	8.1	13.3	0.99	11.1	8.4	16.2	0.95
30	31.5	24.3	35.8	0.97	11.0	9.2	13.5	1.00	12.8	9.6	17.0	1.08
35	31.9	27.5	36.8	0.98	11.2	8.8	12.4	1.01	10.6	9.2	13.6	0.91
40	34.1	31.1	36.0	1.05	11.2	8.5	13.3	1.02	11.5	8.8	16.0	0.99
45	32.9	30.9	36.4	1.01	10.6	8.5	12.5	0.96	12.3	10.3	14.1	1.05
50	31.2	27.2	35.6	0.96	11.0	9.4	12.2	0.99	11.7	10.0	13.1	1.00
100	32.4				11.1				11.7			

Generation	Number of plots
F ₃	164
F ₄	261
F ₅	265

Low: lowest within plot variance; Hig: highest within plot variance; $\overline{\sigma_d^2}$: sample of 100%; σ_d^2 : sample of *i*%; *i* = 5, 10, ..., 50.

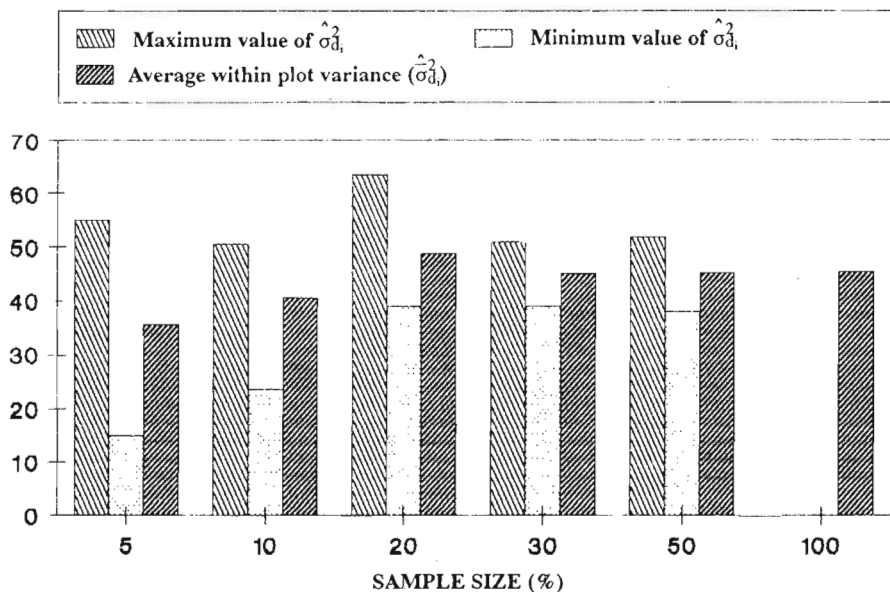


Figure 1 - Estimate of the average within plot variance ($\hat{\sigma}_d^2$) of the number of pods per plant in the F_3 generation.

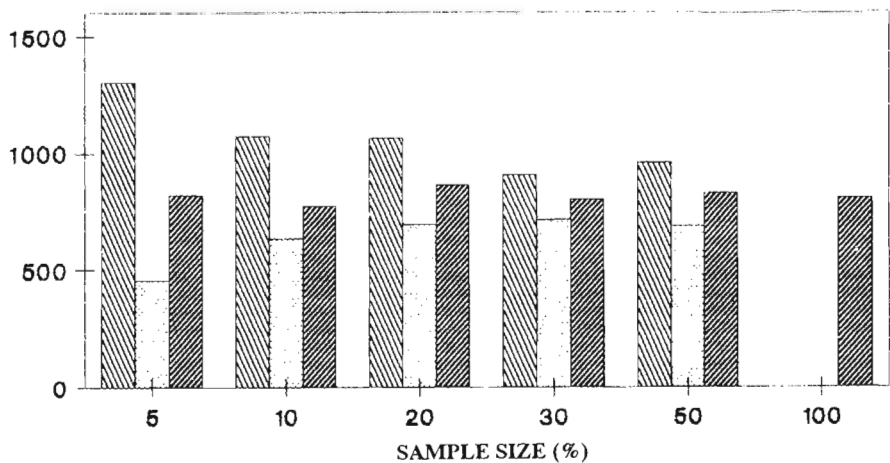


Figure 2 - Estimate of the average within plot variance ($\hat{\sigma}_d^2$) of the number of seeds per plant in the F_3 generation.

plant level in a reduced number of plots. This procedure, besides reducing the work will also diminish the error margin in the values obtained, because if all the plots were to be considered the number of plants handled would be very large, increasing note-taking errors, plant loss and other problems.

The choice of sample size will depend on several factors, such as available resources and the number of plots in the experiments. In similar situations, which involve experiments with over 150 plots, a sample of 15% to 20%, corresponding to 23 to 30 plots, would be adequate. Besides supplying a calculation of the average $\hat{\sigma}_d^2$ similar to the other sample sizes, it presents a range of variation comparable to that observed when 50% of the plots are sampled. If the total number of plots is smaller, larger samples should be chosen, and the opposite strategy should be taken if the number of plots is very large.

RESUMO

Com o objetivo de verificar o uso de amostragem de parcelas para quantificar a variação dentro das progêneses de feijoeiro e determinar o tamanho mínimo da amostra para que seja representativa foram avaliadas 93 progêneses segregantes do cruzamento Jalo x Small White nas gerações F₃, F₄ e F₅ em experimentos conduzidos durante o período das secas de 1991. Nesses experimentos foram coletadas individualmente cinco plantas por parcela e estimada a média e a variância da produção de grãos, número de vagens, e o número de sementes por planta. Foram simulados diferentes tamanhos amostrais do número de parcelas, variando de 5 a 50%. Em cada simulação foram retiradas dez amostras de mesmo tamanho. Comparando as estimativas obtidas por amostra com as da população constatou-se que as amostras representaram bem a população mesmo quando se utilizou o menor tamanho, ou seja, de 5% das parcelas e para todas as características, indiferentemente da geração.

REFERENCES

- Fehr, W.R. (1987). *Principles of Cultivar Development Theory and Technique*. MacMillan Pub. Co., New York, pp. 748.
- Pacheco, C.A.P. (1987). Avaliação de famílias de meios irmãos da população de milho CMS-39 em diferentes condições de ambiente. 2º ciclo de seleção. Master's Thesis, ESAL, Lavras.
- Ramalho, M.A.P. and Vencovsky, R. (1978). Estimação dos componentes da variância genética em plantas autógamas. *Ciência e Prática* 2: 117-140.
- Vieira, C. (1967). *O Feijoeiro Comum: Cultura, Doenças e Melhoramento*. Imprensa Universitária, Viçosa, pp. 220.

(Received May 11, 1992)