Sample size for estimating average productive traits of pigeon pea

Tamanho de amostra para a estimação da média de caracteres produtivos de feijão guandu

Giovani Facco¹ Alberto Cargnelutti Filho II Bruna Mendonça Alves¹ Cláudia Burin¹ Gustavo Oliveira dos Santos III Jéssica Andiara Kleinpaul IV Ismael Mario Márcio Neu IV

ABSTRACT

The objectives of this study were to determine the sample size, in terms of number of plants, needed to estimate the average values of productive traits of the pigeon pea and to determine whether the sample size needed varies between traits and between crop years. Separate uniformity trials were conducted in 2011/2012 and 2012/2013. In each trial, 360 plants were demarcated, and the fresh and dry masses of roots, stems, and leaves and of shoots and the total plant were evaluated during blossoming for 10 productive traits. Descriptive statistics were calculated, normality and randomness were checked, and the sample size was calculated. There was variability in the sample size between the productive traits and crop years of the pigeon pea culture. To estimate the averages of the productive traits with a 20% maximum estimation error and 95% confidence level, 70 plants are sufficient.

Key words: Cajanus cajan L., sample size, experimental precision.

RESUMO

Os objetivos deste trabalho foram determinar o tamanho de amostra, em número de plantas, para a estimação da média de caracteres produtivos de feijão guandu e verificar se há variabilidade do tamanho de amostra entre caracteres e anos agrícolas. Foram conduzidos dois ensaios de uniformidade, um em 2011/2012 e o outro em 2012/2013. Em cada ensaio, foram demarcadas 360 plantas e, no florescimento, foram avaliadas as massas verde e seca de raiz, caule, folha, parte aérea e total, totalizando dez caracteres produtivos. Foram calculadas estatisticas descritivas, verificada a normalidade e a aleatoriedade e calculado o tamanho de amostra. Na cultura de feijão guandu, há variabilidade do tamanho de amostra entre os caracteres produtivos e entre os anos agrícolas. Para a estimação da média dos caracteres produtivos, com erro de estimação máximo de 20% da média e com grau de confiança de 95%, 70 plantas são suficientes.

Palavras-chave: Cajanus cajan L., dimensionamento amostral, precisão experimental.

INTRODUCTION

The pigeon pea (Cajanus cajan (L.) Millsp.) is a legume used in green manure that produces fresh and dry material for soil coverage (SUZUKI & ALVES, 2006; FINHOLDT et al., 2009; CAVALCANTE et al., 2012) used in the recovery of degraded areas (RAYOL & ALVINO-RAYOL, 2012) and as livestock feed (ALENCAR et al., 2014).

The assessment of a large number of traits in all individuals of a population may be unfeasible in agricultural experiments due to limited personnel, financial and time resources. Thus, sampling techniques offer advantages, as they enable the estimation of parameters of plant population to the desired precision (BUSSAB & MORETTIN, 2011). Appropriate sample size for each agricultural crop should be investigated, and is dependent on the variability of the traits being estimated and the researcher’s desired precision. Larger sample sizes provide higher experimental precision since they decrease the variance of the
seeds were sown at a density of 20 seeds m⁻² year⁻¹ (2012/2013), in the same 28m×66m (1,848m²) area. On 20 November 2012, during the second crop year, 20 seeds m⁻² was carried out in a 28m×66m (1,848m²) area. The first trial was carried out in the 2011/2012 crop year, and the second in 2012/2013. Both trials were conducted uniformly across the experimental area. The first trial was carried out in the 2011/2012 crop year, and the second in 2012/2013. Both trials were conducted in the same experimental area of the Department of Plant Science of the Universidade Federal de Santa Maria, located 30m from the experimental area. The average daily air temperature, in °C, was calculated as the arithmetic mean of the minimum and maximum air temperatures.

The following statistics were calculated for each of the 10 traits evaluated in each crop year: minimum (min), maximum (max), mean (x̄), median (med), standard deviation (s), standard error of the mean (se), coefficient of variation (CV, as a percentage), asymmetry (A), and kurtosis (K). Kolmogorov-Smirnov tests and run tests were conducted to verify the assumptions of normality and randomness, respectively. For the run test, based on the median, a layout was considered starting in line 1, column 1, going through to the line 1, column 12, returning to start at line 2, column 12, going up to line 2, column 1, and starting over in line 3, column 1, up to line 3, column 12, and so on, until reaching line 30, column 1. The average of the RFM, SFM, LFM, FMOS, TFM, RDM, SDM, LDM, DMOS, and TDM traits between both crop years (n=360 plants per crop year) were compared using a Student’s t-test for independent samples (P=0.05, df=718).
Subsequently, for each trait in each crop year, the sample size (n) was calculated for the semi-amplitude of the confidence interval (estimation error), set at 2%, 4%, 6%, 8%, 10%, 12%, 14%, 16%, 18%, and 20% of the average (\(\bar{x}\)) of each trait, that is, from 0.02 \(\bar{x}\) (high precision) up to 0.20 \(\bar{x}\) (low precision), with a 95% confidence level (1-\(\alpha\)). The expression \(n = [(t_{\alpha/2} s)/(\text{estimation error})]^2\) was used (BUSSAB & MORETTIN, 2011), in which \(t_{\alpha/2}\) is the critical value of the Student’s \(t\) distribution, such that \(P(t> t_{\alpha/2})=\alpha/2\), with a 5% probability of error and n-1 degrees of freedom (359 degrees of freedom in this study), where \(s\) is the estimate of the standard deviation.

From the equation used to calculate the sample size, \(n\) was set at 360 plants, which was the number of sampled plants each year, and the estimation error (semi-amplitude of the confidence interval) was calculated as a percentage of the estimated average (\(\bar{x}\)), for each trait, using the equation: estimation error = 100 \((t_{\alpha/2} s)/(\sqrt{n} \bar{x})\). The statistical analyses were conducted using the GENES application (CRUZ, 2013) and Microsoft Office Excel®.

RESULTS AND DISCUSSION

The minimum and maximum values of root fresh mass (RFM), stem fresh mass (SFM), leaf fresh mass (LFM), fresh mass of shoots (FMOS), total fresh mass (TFM), root dry mass (RDM), stem dry mass (SDM), leaf dry mass (LDM), dry mass of shoots (DMOS), and total dry mass (TDM) traits in grams per plant (Table 1), in both crop years, showed variability among the sampled plants. The variability among the 360 plants in each crop year, and between crop years, is important in determining the appropriate sample size, as it reflects real situations in field conditions. In addition to environmental variations between crop years (Figure 1A and 1B), this variability is likely associated with the fact that distinct sowing systems were used (broadcast and row seeding), and the different durations for which plants were in the field prior to evaluation (125 and 162 DAS).

In the 2011/2012 crop year, the 360 plants sampled had an average RFM (72.53g·plant\(^{-1}\)) and LFM (130.71g·plant\(^{-1}\)) higher than the average RFM (49.63g·plant\(^{-1}\)) and LFM (94.02g·plant\(^{-1}\)) obtained in the 360 plants sampled in the 2012/2013 crop year. The SFM showed the opposite behavior, i.e., the average in the 2011/2012 crop year (235.27g·plant\(^{-1}\)) was lower than that in the 2012/2013 crop year (266.51g·plant\(^{-1}\)) (Table 1). This change in behavior possibly explains the fact that the average of the fresh mass of shoots (FMOS=SFM+LFM) and the total fresh mass (TFM=RFM+SFM+LFM) did not differ between crop years. Consequently, as expected, the traits related to the dry mass (RDM, SDM, LDM, DMOS, and TDM) exhibited the same behavior between the crop years as the traits related to the fresh mass (RFM, SFM, LFM, FMOS, and TFM).

On average between the two crop years, the root dry mass (RDM), stem dry mass (SDM), leaf dry mass (LDM), dry mass of shoots (DMOS), and total dry mass (TDM) accounted for 31.83% of RFM, 29.87% of SFM, 27.21% of LFM, 29.13% of FMOS, and 29.51% of TFM, respectively. Studies by SUZUKI & ALVES (2006) and FINHOLDT et al. (2009) found pigeon pea dry masses of 19.8% and 16.8% of the fresh masses, respectively. Conversely, CAVALCANTE et al. (2012) found percentages closer to those obtained in this study, that is, 26.0% for dwarf pigeon pea and 29.5% for forage pigeon pea. These dry mass percentages convey the important contributions of this culture as a cover crop.

In the 2011/2012 and 2012/2013 crop years, the traits related to the fresh mass (RFM, SFM, LFM, FMOS, and TFM) had standard deviations that varied between 41.28 and 288.26g·plant\(^{-1}\), standard errors that varied between 2.18 and 15.19g·plant\(^{-1}\), and coefficients of variation (CVs) that varied between 55.36 and 83.18% (Table 1). For the dry mass traits (RDM, SDM, LDM, DMOS, and TDM), the standard deviations varied between 13.58 and 88.85g·plant\(^{-1}\), the standard errors varied between 0.72 and 4.68g·plant\(^{-1}\), and the CVs between 16.8% and 84.53%. In general, among the traits in each crop year, higher CVs were obtained for the RFM and RDM than for the other traits. This difference was more significant in the 2012/2013 crop year, with a CV of 83.18% for RFM and 84.53% for RDM, while the CVs of the other eight traits varied between 69.47% and 73.03% (Table 1). In the second crop year (row seeding, and evaluation at 162DAS), the productive traits of the fresh and dry pigeon pea masses exhibited higher variability (mean CV of the 10 traits = 73.77%) than in the first crop year (broadcast seeding and evaluation at 125DAS; mean CV of the 10 traits = 58.73%). This suggests that to estimate the averages of the productive traits of pigeon pea with the same precision, the sample size needs to be different between traits and between crop years. Coefficients of variation for productive
traits with a similar magnitude were also obtained in the sample size studies with cover crops such as white lupine (47.58≤CV≤57.58%) (BURIN et al., 2014), forage turnip (62.47≤CV≤77.07%) (CARGNELUTTI FILHO et al., 2014), and black oat (31.58≤CV≤67.96%) (CARGNELUTTI FILHO et al., 2015).

For 10 traits, in both crop years, the high mean scores in relation to the median, the positive and non-zero estimates of asymmetry, and kurtosis values other than 3, as well as the significance of the Kolmogorov-Smirnov test (P≤0.05), indicate that the data do not fit the normal distribution. According to the randomness test (run test), in 70% of the cases (10 traits × 2 crop years) the distribution was random (P>0.05) (Table 1). Although the assumption of normality was not satisfied, the large number of sampled plants (360 plants in each crop years) and the high percentage of cases with randomness associated with the broad variability between agricultural traits and crop years, gives credibility to the use of this database for the sample size study (SIEGEL & CASTELLAN JR, 2006; BUSSAB & MORENTTIN, 2011).

The sample size, in number of plants, for estimating the average of the 10 productive traits in both crop years, with an estimation error equal to 2% from the estimated average (X) and a 95% confidence level, varied between 2,963 plants for LFM (first crop year) and 6,909 plants for RDM (second crop year) (Table 2). In general, in each crop year, larger sample sizes were needed for RFM and RDM than for the other traits (Table 2). In the second crop year, the sample sizes needed for estimating the average of the productive traits of pigeon pea fresh and dry masses were higher than in the first crop year.

Table 1 - Minimum (min), maximum (max), mean, median (med), standard deviation (s), standard error (se), coefficient of variation (CV, in %), asymmetry (A), kurtosis (K), Kolmogorov-Smirnov normality value test (KS), and P-value of the run test (RT, a randomness test) for root fresh mass (RFM), stem fresh mass (SFM), leaf fresh mass (LFM), fresh mass of shoots (FMOS), total fresh mass (TFM), root dry mass (RDM), stem dry mass (SDM), leaf dry mass (LDM), dry mass of shoots (DMOS), and total dry mass (TDM), in g plant−1, of 360 pigeon pea (Cajanus cajan L.) plants evaluated at 125 days after sowing in the 2011/2012 crop year, and of 360 plants evaluated at 162 days after sowing in the 2012/2013 crop year.

<table>
<thead>
<tr>
<th>Trait</th>
<th>min</th>
<th>max</th>
<th>mean</th>
<th>med(1)</th>
<th>s</th>
<th>se</th>
<th>CV (%)</th>
<th>A(2)</th>
<th>K(3)</th>
<th>KS(4)</th>
<th>RT(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RFM</td>
<td>9.00</td>
<td>237.00</td>
<td>72.53</td>
<td>a</td>
<td>65.00</td>
<td>45.21</td>
<td>62.34</td>
<td>1.01</td>
<td>3.90</td>
<td>1.70</td>
<td>0.01</td>
</tr>
<tr>
<td>SFM</td>
<td>33.00</td>
<td>746.00</td>
<td>235.27</td>
<td>b</td>
<td>206.30</td>
<td>139.04</td>
<td>59.10</td>
<td>1.09</td>
<td>4.22</td>
<td>1.73</td>
<td>0.02</td>
</tr>
<tr>
<td>LFM</td>
<td>9.00</td>
<td>1,128.00</td>
<td>365.98</td>
<td>a</td>
<td>325.00</td>
<td>209.86</td>
<td>57.34</td>
<td>1.10</td>
<td>4.27</td>
<td>1.58</td>
<td>0.03</td>
</tr>
<tr>
<td>FMOS</td>
<td>3.00</td>
<td>75.00</td>
<td>22.70</td>
<td>a</td>
<td>20.50</td>
<td>14.00</td>
<td>61.69</td>
<td>1.00</td>
<td>3.83</td>
<td>1.81</td>
<td>0.01</td>
</tr>
<tr>
<td>SDM</td>
<td>6.00</td>
<td>227.00</td>
<td>68.13</td>
<td>b</td>
<td>50.95</td>
<td>40.95</td>
<td>60.11</td>
<td>1.13</td>
<td>4.39</td>
<td>2.15</td>
<td>0.07</td>
</tr>
<tr>
<td>LDM</td>
<td>13.00</td>
<td>335.00</td>
<td>127.50</td>
<td>a</td>
<td>111.00</td>
<td>74.68</td>
<td>58.57</td>
<td>1.08</td>
<td>4.23</td>
<td>1.77</td>
<td>0.25</td>
</tr>
<tr>
<td>DMOS</td>
<td>13.00</td>
<td>400.00</td>
<td>127.50</td>
<td>a</td>
<td>111.00</td>
<td>74.68</td>
<td>58.57</td>
<td>1.08</td>
<td>4.23</td>
<td>1.77</td>
<td>0.25</td>
</tr>
</tbody>
</table>

(1)For each trait (RFM, SFM, LFM, FMOS, TFM, RDM, SDM, LDM, DMOS, and TDM) averages not followed by the same letter in the column (mean comparison between crop years) are different at a 5% level of probability by the Student’s t-test for independent samples, with 718 degrees of freedom. (2)*Non-zero asymmetry, by t-test, at a 5% level of probability. (3)*Kurtosis other than three, by t-test, at a 5% level of probability. (4)*Significant at 5% probability (normal distribution). (5)*P-value = 0.05 (non-random data series) and P-value > 0.05 (random data series).
Thus, it can be inferred that, in order to estimate the average of the productive traits of pigeon pea with the same precision, different sample sizes are required for different traits and crop years.

Sample size variability between seed, seedling, and adult plants for castor bean production traits was observed by CARGNELUTTI FILHO et al. (2010). In addition, sample size variability between

Ciência Rural, v.46, n.4, abr, 2016.
morphological and productive traits was observed in cover crops, such as white lupine (BURIN et al., 2014), forage turnip (CARGNELUTTI FILHO et al., 2014), and black oat (CARGNELUTTI FILHO et al., 2015). Additionally, sample size variability between evaluation periods was observed for the black oat (CARGNELUTTI FILHO, 2015).

In practice, estimating the average of these productive traits of pigeon pea with high precision, i.e., with a 2% maximum estimation error, requires an unfeasibly large number of plants (6,909 plants, Table 2). Alternatively, smaller sample sizes may be used, but these are associated with higher estimation errors, so the estimate of the average will have lower precision. With a 2% decrease in precision, that is, with an estimation error of 4%, there is a 75% reduction in the sample size for all traits in both crop years. For the semi-amplitude of the confidence interval equal to 20% of the average (lower precision, in this study), at least 70 plants are required for the productive traits of fresh and dry mass for both crop years (Table 2). In practice, in sample survey studies, these 70 plants would be taken at random from the experimental area. In an experimental design with, for instance, five replicates per treatment, 14 plants per repetition would be sampled.

Researchers can use the data obtained from this study for the planning of appropriate samples for pigeon pea research. The traits being evaluated, the desired precision, the availability of personnel, financial resources, and time are important aspects to be considered in the experimental design. In addition to the approach suggested by this study, one can, if needed, estimate the sample size for other estimation errors.

CONCLUSION

The necessary sample size varies with the productive traits being assessed and the crop years of the pigeon pea. To estimate the averages of the productive traits with a 20% maximum estimation error and 95% confidence level, 70 plants are sufficient.

ACKNOWLEDGEMENTS

To the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), for granting...
scholarships to the authors. To the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), for the financial support.

REFERENCES


