CARGNELUTTI FILHO, A.; SILVEIRA, D. L.; BUBANS, V. E.; OSMARI, B. F.; SOMAVILLA, F. M.; ORTIZ, V. M.; OSMARI, L. F. Sample size and linear relations in slender leaf rattlebox. **Revista de Agricultura Neotropical**, Cassilândia-MS, v. 9, n. 2, e6834, abr./maio. 2022. ISSN 2358-6303. DOI: https://doi.org/10.32404/rean.v9i2.6834.

Sample size and linear relations in slender leaf rattlebox

Alberto Cargnelutti Filho[1], Daniela Lixinski Silveira[1], Valéria Escaio Bubans[1], Bruno Fillipin Osmari[1], Felipe Manfio Somavilla[1], Vithória Morena Ortiz[1], Lucas Fillipin Osmari[1]

[1] Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, Brasil. E-mail: alberto.cargnelutti.filho@gmail.com, danilisil@gmail.com, valeriabubans@hotmail.com, brunoosmari11@gmail.com, felipe_somavilla@hotmail.com, vithoria.ortiz159@gmail.com, lucasfosmari@gmail.com

Received: 12/01/2022; Accepted: 18/04/2022.

ABSTRACT

The objectives of this work were to determine the sample size (number of plants) to estimate the mean of slender leaf rattlebox (*Crotalaria ochroleuca*) traits and investigate the relations among traits. In an experimental area of 8 m × 20 m (160 m^2), at 147 days after sowing, 110 plants were randomly selected. The following traits were evaluated for each plant: plant height, stem diameter, number of nodes, number of leaves, leaf fresh matter, stem fresh matter, shoot fresh matter, leaf dry matter, stem dry matter and shoot dry matter. Measures of central tendency, variability, skewness and kurtosis were calculated. Normality was verified using the Kolmogorov-Smirnov test and the sample size was calculated to estimate the mean of the traits, assuming estimation errors (semi-amplitudes of the 95% confidence interval) equal to 1%, 2%, …, 30% of the mean. The relations among traits were investigated by scatter plots, correlation and path analysis. To estimate the mean of these ten traits, with a maximum error of 10% of the mean and 95% confidence level, 67 plants are needed. In an experiment, to estimate the mean of each treatment with 10% precision, 67 plants per treatment must be evaluated. The number of leaves has a positive linear relation with leaf fresh and dry matter. Number of leaves and stem diameter have a positive linear relation with stem and shoot fresh and dry matter.

Keywords: *Crotalaria ochroleuca*, sample sizing, experimental precision, correlation, path analysis.

Tamanho de amostra e relações lineares em crotalária ochroleuca

RESUMO

Os objetivos deste trabalho foram determinar o tamanho de amostra (número de plantas) necessário para a estimação da média de caracteres de crotalária ochroleuca (*Crotalaria ochroleuca*) e investigar as relações entre os caracteres. Em área experimental de 8 m × 20 m (160 m^2), aos 147 dias após a semeadura, foram selecionadas, aleatoriamente, 110 plantas. Em cada planta foram avaliados os seguintes caracteres: altura de planta, diâmetro de caule, número de nós, número de folhas, matéria fresca de folhas, matéria fresca de caule, matéria fresca de parte aérea, matéria seca de folhas, matéria seca de caule e matéria seca de parte aérea. Foram calculadas as medidas de tendência central, variabilidade, assimetria e curtose. Foi verificada a normalidade por meio do teste de Kolmogorov-Smirnov e calculado o tamanho de amostra para a estimação da média dos caracteres, assumindo erros de estimação (semi-amplitudes do intervalo de confiança de 95%) iguais a 1%, 2%, …, 30% da média. Foi investigada a relação entre os caracteres por meio de diagramas de dispersão, análises de correlação e de trilha. Para a estimação da média desses dez caracteres, com erro máximo de 10% da média e grau de confiança de 95%, são necessárias 67 plantas. Em um experimento, para a estimação da média de cada tratamento com 10% de precisão, devem ser avaliadas 67 plantas por tratamento. O número de folhas tem relação linear positiva com as matérias fresca e seca de folhas. O número de folhas e o diâmetro de caule têm relação linear positiva com as matérias fresca e seca de caule e de parte aérea.

Palavras-chave: *Crotalaria ochroleuca*, dimensionamento amostral, precisão experimental, correlação, análise de trilha.

1. Introduction

In field experiments, with different agricultural crops, measuring several plant traits is important to make use of the resources invested in the research. Measuring in all plants (individuals) of the usable area of the experimental unit (plot) is adequate to represent the trait under evaluation. To minimize labor, time and financial and human resources, it is common to measure part of the plants (sample) in the experimental unit, and the sample should be representative of the plants in the experimental unit (Storck et al., 2016).

Sample size is directly proportional to the variability of the data and to the desired confidence level in the estimate and inversely proportional to the estimation error allowed (Ferreira, 2009; Bussab and Morettin, 2017). Thus, it is important to adequately size the number of plants that must be evaluated to allow precise inferences about the traits under evaluation. Sample sizing, to estimate the mean of traits, has been determined for traits of species of the Fabaceae family, such as: *Lupinus albus* (Burin et al., 2014), *Cajanus cajan* (Facco et al., 2015; Facco et al., 2016), *Crotalaria juncea* (Teodoro et al., 2015; Schabarum et al., 2018a; Schabarum et al., 2018b; Toebe et al., 2018), *Crotalaria spectabilis* (Teodoro et al., 2015; Toebe et al., 2017; Toebe et al., 2018), *Crotalaria breviflora*, *Crotalaria ochroleuca* (Toebe et al., 2018) and *Canavalia ensiformis* (Cargnelutti Filho et al., 2018a). These studies have reported variation in sample size between traits and species.

In addition to individual studies of traits, their interrelations provide important information, especially in the identification of traits for indirect selection. Linear relations among a set of traits can be visualized in scatter plots and investigated using Pearson's linear correlation coefficient (r) and complementary statistical procedures, such as path analysis. The coefficient r is a statistic used to measure the intensity or degree of linear relation between two random variables. The r sign expresses the direction of association, and the intensity is represented by a numerical value between -1 and 1. In extreme situations, two traits may have perfect negative (r = -1) or perfect positive (r = 1) linear correlation, or lack of linear relation (r = 0) (Ferreira, 2009; Bussab and Morettin, 2017).

Path analysis makes it possible to decompose the correlation coefficients into direct and indirect effects of explanatory variables on a main variable and to identify if there is linear association of cause and effect, enabling the identification of traits for indirect selection (Cruz et al., 2012; Cruz et al., 2014). These statistical procedures have been used to investigate the linear association among traits of *Raphanus sativus* and *Lupinus albus* (Cargnelutti Filho et al., 2014), *Crotalaria spectabilis* (Toebe et al., 2017), *Cajanus cajan* (Cargnelutti Filho et al., 2017) and *Canavalia ensiformis* (Cargnelutti Filho et al., 2018b).

The inclusion of studies on sample sizing and association among slender leaf rattlebox (*C. ochroleuca*) traits adds important information to support the planning of experiments with better precision and, consequently, with greater reliability of the results. Thus, the objectives of this work were to determine the sample size (number of plants) necessary to estimate the mean of slender leaf rattlebox traits and investigate the relations among traits.

2. Material and Methods

A uniformity trial (blank experiment) was conducted with slender leaf rattlebox (*C. ochroleuca*), common cultivar, in Santa Maria, Rio Grande do Sul state, Brazil (29º42'S latitude, 53º49'W longitude and 95 m altitude), in an 8 m × 20 m area (160 m²). According to the Köppen-Geiger classification, the climate of this site is humid subtropical - Cfa, with hot summers and no dry season (Alvares et al., 2013). The soil is *Argissolo Vermelho Distrófico Arênico* (Ultisol) (Santos et al., 2018) of sandy loam surface texture. On October 28, 2020, 35 kg ha^{-1} of N, 135 kg ha^{-1} of P_2O_5 and 135 kg ha^{-1} of K_2O were incorporated into the soil, and sowing was carried out broadcast using 18.75 kg of seeds ha^{-1}.

In the flowering of the crop, that is, at 147 days after sowing, 110 plants were randomly selected. In each plant, the following traits were evaluated: plant height (PH, in m), stem diameter (SD, in cm, measured with a caliper at 2 cm from the soil surface), number of nodes (NN), number of leaves (NL), leaf fresh matter (LFM, in g plant^{-1}), stem fresh matter (STFM, in g plant^{-1}), shoot fresh matter (SHFM = LFM + STFM, in g plant^{-1}), leaf dry matter (LDM, in g plant^{-1}), stem dry matter (STDM, in g plant^{-1}) and shoot dry matter (SHDM = LDM + STDM, in g plant^{-1}).

For the ten traits evaluated (PH, SD, NN, NL, LFM, STFM, SHFM, LDM, STDM and SHDM), the following statistics were calculated: minimum, percentiles 1, 2.5 and 25, median (percentile 50), percentiles 75, 97.5 and 99, maximum, amplitude, mean, variance, standard deviation, standard error, coefficient of variation, skewness, kurtosis and p-value of the Kolmogorov-Smirnov normality test. For each trait, based on the 110 plants, the sample size (n) was calculated for the estimation errors (semi-amplitudes of the confidence interval) fixed at 1%, 2%, ..., 30% of the mean (m), that is, 0.01×m (higher precision), 0.02×m, ... 0.30×m (lower precision), with a 95% confidence level (1-α) by the expression (Ferreira, 2009; Bussab and Morettin, 2017):

$$n = \left(\frac{t_{\alpha/2}\, s}{estimation\ error}\right)^2$$

Where $t_{\alpha/2}$ is the critical value of the Student's t-distribution, whose area on the right is equal to α/2, that is, the value of t, such that $P(t>t_{\alpha/2})=\alpha/2$, with α=5% probability of error and n-1 degrees of freedom (109 degrees of freedom in this study), and s is the estimate of the standard deviation. After, fixing n equal to 110 plants, which was the sample size used in the sampling, the estimation error was calculated as a percentage of the mean estimate (m), for each of the traits, using the expression:

$$estimation\ error = \frac{t_{\alpha/2}\, s}{\sqrt{n}\, m} \times 100$$

To study relations among the traits PH, SD, NN, NL, LFM, STFM, SHFM, LDM, STDM and SHDM, scatter plots were initially constructed. Then, the matrix of Pearson's linear correlation coefficients (r) among the traits was determined and the significance of r was checked using Student's t-test, at 5% probability of error. In the correlation matrix among the traits PH, SD, NN and NL, multicollinearity was diagnosed (Cruz et al., 2014) and interpreted based on the condition number (CN), which is the ratio between the highest and lowest eigenvalue of the correlation matrix. Multicollinearity was classified as weak when CN ≤ 100, moderate to severe when 100 < CN < 1000 and severe when CN ≥ 1000, according to Montgomery and Peck´s criteria (1982).

After the diagnosis of multicollinearity, path analysis of the main variables (LFM, STFM, SHFM, LDM, STDM, SHDM) as a function of the explanatory variables (PH, SD, NN and NL) was carried out according to the methodology described in Cruz et al. (2012) and Cruz et al. (2014), totaling six path analyses. For example, the first path analysis considered LFM as a function of PH, SD, NN and NL, the second path analysis considered STFM as a function of PH, SD, NN and NL, and so on, up to the sixth path analysis, which considered SHDM as a function of PH, SD, NN and NL. Microsoft Office Excel® and Genes (Cruz, 2016) applications were used to perform the statistical analyses.

3. Results and Discussion

In relation to plant height (PH), stem diameter (SD), number of nodes (NN), number of leaves (NL), leaf fresh matter (LFM), stem fresh matter (STFM), shoot fresh matter (SHFM), leaf dry matter (LDM), stem dry matter (STDM) and shoot dry matter (SHDM), the p-value of the Kolmogorov-Smirnov test ranged from 0.11 to 0.61 (Table 1). The higher the p-value, the greater the adherence of the data to the normal distribution curve. Thus, assuming the significance level of 11%, it is possible to infer that the assumption of normality was met for all traits. Coefficients of skewness and kurtosis close to zero and three, respectively, although significant for some traits, and proximity of the mean to the median, reinforce the fit of the data to the normal distribution. Therefore, this data set is suitable for the studies of sample sizing based on Student's t-distribution and relations by means of correlation and path analyses.

Table 1. Minimum, percentiles 1, 2.5 and 25, median (percentile 50), percentiles 75, 97.5 and 99, maximum, amplitude, mean, variance, standard deviation, standard error, coefficient of variation (CV), skewness, kurtosis and p-value of the Kolmogorov-Smirnov normality test of traits[(1)] measured in 110 slender leaf rattlebox (*C. ochroleuca*) plants.

Statistic	PH	SD	NN	NL	LFM	STFM	SHFM	LDM	STDM	SHDM
Minimum	2.27	1.10	58.00	71.00	27.50	151.50	179.00	3.50	36.50	40.00
Percentile 1	2.43	1.11	62.27	84.45	36.05	154.45	198.77	4.05	38.59	44.05
Percentile 2.5	2.50	1.20	65.73	89.00	38.31	165.16	209.04	4.86	39.86	46.31
Percentile 25	2.93	1.60	82.00	155.00	59.75	269.63	331.75	10.00	69.00	80.13
Median	3.13	1.70	88.00	222.00	78.75	315.75	400.25	12.50	84.50	97.75
Percentile 75	3.34	1.90	93.00	295.75	96.25	397.13	485.25	16.00	104.00	117.50
Percentile 97.5	3.57	2.20	105.00	467.53	148.86	562.31	698.13	27.05	144.95	171.99
Percentile 99	3.66	2.20	105.91	522.76	167.10	593.66	767.12	30.32	167.56	197.52
Maximum	3.66	2.30	107.00	558.00	174.50	628.00	795.50	31.50	185.00	216.50
Amplitude	1.39	1.20	49.00	487.00	147.00	476.50	616.50	28.00	148.50	176.50
Mean	3.10	1.73	87.12	233.45	81.09	333.57	414.66	13.37	87.94	101.30
Variance	0.09	0.06	106.95	9274.74	794.39	9867.08	15347.97	29.85	798.07	1094.89
Standard deviation	0.30	0.25	10.34	96.31	28.18	99.33	123.89	5.46	28.25	33.09
Standard error	0.03	0.02	0.99	9.18	2.69	9.47	11.81	0.52	2.69	3.15
CV (%)	9.68	14.59	11.87	41.25	34.76	29.78	29.88	40.87	32.13	32.66
Skewness [(2)]	-0.51*	-0.07ns	-0.43ns	0.91*	0.94*	0.55*	0.63*	0.97*	0.69*	0.74*
Kurtosis + 3 [(3)]	2.76ns	2.93ns	2.96ns	4.00*	4.14*	3.21ns	3.52ns	4.29*	3.75ns	3.92*
p-value	0.53	0.17	0.61	0.23	0.11	0.45	0.41	0.15	0.40	0.27

[(1)] PH - plant height, in m; SD - stem diameter, in cm; NN - number of nodes; NL - number of leaves; LFM - leaf fresh matter, in g $plant^{-1}$; STFM - stem fresh matter, in g $plant^{-1}$; SHFM - shoot fresh matter (SHFM = LFM + STFM), in g $plant^{-1}$; LDM - leaf dry matter, in g $plant^{-1}$; STDM - stem dry matter, in g $plant^{-1}$; and SHDM - shoot dry matter (SHDM = LDM + STDM), in g $plant^{-1}$. [(2)] * Skewness differs from zero, through Student's t-test, at 5% probability level. ns Not significant. [(3)] * Kurtosis differs from three, through the Student's t-test, at 5% probability level. ns Not significant.

Based on the percentiles and dispersion measures, there was a wide variation among the 110 plants for all traits (Table 1). This wide variation is important for the studies of sample sizing and relations through correlation and path analyses, because it contemplates plants of different sizes (small, medium and large), which are common in field experiments.

The coefficient of variation (CV), in relation to the ten traits (PH, SD, NN, NL, LFM, STFM, SHFM, LDM, STDM and SHDM), ranged between 9.68% for PH and 41.25% for NL, with mean of 27.75% (Table 1). The coefficients of variation of the traits NL, LFM, STFM, SHFM, LDM, STDM and SHDM (29.78% ≤ CV ≤ 41.25%) were higher than those of PH, SD and NN (9.68% ≤ CV ≤ 14.59%). Thus, for the same precision, a larger sample size is expected to estimate the mean of the traits NL, LFM, STFM, SHFM, LDM, STDM and SHDM, compared to the traits PH, SD and NN. The sample sizes (number of plants) for estimating the mean, with estimation error (semi-amplitude of the 95% confidence interval) equal to 1% of the mean estimate (m), that is, 0.01×m (higher precision, in this study), ranged from 369 plants for PH to 6686 plants for NL (Table 2).

Therefore, in relation to plant height, it can be inferred, with 95% confidence, that the confidence interval of the mean (μ), obtained with 369 plants, is μ±0.01m, that is, μ±0.031 meters, because the mean height of the 110 plants sampled was 3.10 meters (Table 1). In the other extreme, the precision of μ±0.01m is obtained with 6686 plants, for the number of leaves per plant and, in this situation, the value would have been μ±2.33 leaves, because the mean for number of leaves of the 110 plants sampled was 233.45 leaves (Table 1).

Table 2. Sample size (number of plants) for estimating the mean of traits [(1)] of slender leaf rattlebox (*C. ochroleuca*), for estimation errors (semi-amplitudes of the confidence interval) equal to 1%, 2%, ..., 30% of the mean (m), that is, 0.01×m (higher precision), 0.02×m, ... 0.30×m (lower precision), with a 95% confidence level (1-α). Estimation error, as a percentage of the mean estimate (m), based on the 110 plants sampled (Error, in %).

Error	PH	SD	NN	NL	LFM	STFM	SHFM	LDM	STDM	SHDM
1%	369	836	554	6686	4746	3484	3507	6562	4055	4191
2%	93	209	139	1672	1187	871	877	1641	1014	1048
3%	41	93	62	743	528	388	390	730	451	466
4%	24	53	35	418	297	218	220	411	254	262
5%	15	34	23	268	190	140	141	263	163	168
6%	11	24	16	186	132	97	98	183	113	117
7%	8	18	12	137	97	72	72	134	83	86
8%	6	14	9	105	75	55	55	103	64	66
9%	5	11	7	83	59	44	44	82	51	52
10%	4	9	6	67	48	35	36	66	41	42
11%	4	7	5	56	40	29	29	55	34	35
12%	3	6	4	47	33	25	25	46	29	30
13%	3	5	4	40	29	21	21	39	24	25
14%	2	5	3	35	25	18	18	34	21	22
15%	2	4	3	30	22	16	16	30	19	19
16%	2	4	3	27	19	14	14	26	16	17
17%	2	3	2	24	17	13	13	23	15	15
18%	2	3	2	21	15	11	11	21	13	13
19%	2	3	2	19	14	10	10	19	12	12
20%	1	3	2	17	12	9	9	17	11	11
21%	1	2	2	16	11	8	8	15	10	10
22%	1	2	2	14	10	8	8	14	9	9
23%	1	2	2	13	9	7	7	13	8	8
24%	1	2	1	12	9	7	7	12	8	8
25%	1	2	1	11	8	6	6	11	7	7
26%	1	2	1	10	8	6	6	10	6	7
27%	1	2	1	10	7	5	5	10	6	6
28%	1	2	1	9	7	5	5	9	6	6
29%	1	1	1	8	6	5	5	8	5	5
30%	1	1	1	8	6	4	4	8	5	5
Error (%)	1.83	2.76	2.24	7.80	6.57	5.63	5.65	7.72	6.07	6.17

(1) PH - plant height, in m; SD - stem diameter, in cm; NN - number of nodes; NL - number of leaves; LFM - leaf fresh matter, in g $plant^{-1}$; STFM - stem fresh matter, in g $plant^{-1}$; SHFM - shoot fresh matter (SHFM = LFM + STFM), in g $plant^{-1}$; LDM - leaf dry matter, in g $plant^{-1}$; STDM - stem dry matter, in g $plant^{-1}$; and SHDM - shoot dry matter (SHDM = LDM + STDM), in g $plant^{-1}$.

These results reveal that, for the same precision, the sample sizes vary among traits, as observed in species of the Fabaceae family, such as: *Lupinus albus* (Burin et al., 2014), *Cajanus cajan* (Facco et al., 2015; Facco et al., 2016), *Crotalaria juncea*, *Crotalaria spectabilis*, *Crotalaria breviflora*, *Crotalaria ochroleuca* (Teodoro et al., 2015; Schabarum et al., 2018a; Schabarum et al., 2018b; Toebe et al., 2017; Toebe et al., 2018) and *Canavalia ensiformis* (Cargnelutti Filho et al., 2018a). These results also show the difficulty of obtaining estimates of the mean with this higher level of precision due to the high number of plants to be sampled.

With this estimation error of 1%, a smaller sample size was observed for the traits PH, SD and NN ($369 \leq n \leq 836$, with mean of 586 plants) when compared to the sample size required for the traits NL, LFM, STFM, SHFM, LDM, STDM and SHDM ($3484 \leq n \leq 6686$, with mean of 4747 plants) (Table 2). As expected, these results are due to the greater variability of the traits NL, LFM, STFM, SHFM, LDM, STDM and SHDM compared to the traits PH, SD and NN.

If the researcher chooses to allow estimation error of 30%, that is, 0.30×m (lower precision, in this study), and 95% confidence level, the number of plants to be sampled decreases to 1, 1, 1, 8, 6, 4, 4, 8, 5 and 5, respectively, for the traits PH, SD, NN, NL, LFM, STFM, SHFM, LDM, STDM and SHDM (Table 2). It is evident that this low number of plants (≤ 8) would be easily evaluated in an experiment. However, it would lead to low precision in estimating the mean of the traits, that is, μ±0.93 meters, μ±0.52 cm, μ±26.14 nodes, μ±70.03 leaves, μ±24.33 grams, μ±100.07 grams, μ±124.40 grams, μ±4.01 grams, μ±26.38 grams and μ±30.39 grams, respectively, for the traits PH, SD, NN, NL, LFM, STFM, SHFM, LDM, STDM and SHDM.

In practice, the results of this study allow the researcher to choose the sample size, in order to estimate the mean of these traits with the desired precision. For example, if the option is to allow maximum estimation error of 10%, that is, 0.10×m, 67 plants would be sufficient to estimate the mean of these ten traits (Table 2). Thus, when planning an experiment to be conducted in the field, in a completely randomized design to estimate the mean of each treatment with 10% precision, 67 plants per treatment should be evaluated. If the experiment is planned with four replicates per treatment, 17 plants per replicate (67/4 = 16.75) would be sampled, that is, 17 plants per plot. Furthermore, if ten treatments were evaluated in the experiment, 670 plants (67 per treatment) would be sampled.

If the option is to sample 110 plants, as used in this study, the estimation error would be 1.83, 2.76, 2.24, 7.80, 6.57, 5.63, 5.65, 7.72, 6.07 and 6.17% of the mean estimate (m), respectively, for the traits PH, SD, NN, NL, LFM, STFM, SHFM, LDM, STDM and SHDM, that is, a maximum estimation error of 7.80% (Table 2). Pearson's linear correlation coefficients (r) among the traits ranged from 0.171 to 0.997, with mean of 0.661, evidencing positive linear associations among the traits (Table 3). This means that, as the scores of a given trait increase, the scores of another trait increase. This linearity pattern is important for the identification of traits for indirect selection (Figure 1).

Table 3. Estimates of Pearson's linear correlation coefficients among the traits [(1)] measured in 110 plants of slender leaf rattlebox (*C. ochroleuca*).

	PH	SD	NN	NL	LFM	STFM	SHFM	LDM	STDM	SHDM
PH	1.000	0.445*	0.589*	0.171ns	0.212*	0.450*	0.409*	0.189*	0.443*	0.409*
SD	0.445*	1.000	0.346*	0.551*	0.663*	0.847*	0.830*	0.618*	0.796*	0.782*
NN	0.589*	0.346*	1.000	0.290*	0.284*	0.359*	0.353*	0.243*	0.356*	0.344*
NL	0.171ns	0.551*	0.290*	1.000	0.895*	0.762*	0.814*	0.881*	0.792*	0.821*
LFM	0.212*	0.663*	0.284*	0.895*	1.000	0.837*	0.899*	0.969*	0.864*	0.898*
STFM	0.450*	0.847*	0.359*	0.762*	0.837*	1.000	0.992*	0.810*	0.970*	0.962*
SHFM	0.409*	0.830*	0.353*	0.814*	0.899*	0.992*	1.000	0.870*	0.974*	0.976*
LDM	0.189*	0.618*	0.243*	0.881*	0.969*	0.810*	0.870*	1.000	0.865*	0.903*
STDM	0.443*	0.796*	0.356*	0.792*	0.864*	0.970*	0.974*	0.865*	1.000	0.997*
SHDM	0.409*	0.782*	0.344*	0.821*	0.898*	0.962*	0.976*	0.903*	0.997*	1.000

[(1)] PH - plant height, in m; SD - stem diameter, in cm; NN - number of nodes; NL - number of leaves; LFM - leaf fresh matter, in g $plant^{-1}$; STFM - stem fresh matter, in g $plant^{-1}$; SHFM - shoot fresh matter (SHFM = LFM + STFM), in g $plant^{-1}$; LDM - leaf dry matter, in g $plant^{-1}$; STDM - stem dry matter, in g $plant^{-1}$; and SHDM - shoot dry matter (SHDM = LDM + STDM), in g $plant^{-1}$. * Significant at 5% probability of error by Student's t-test, with 108 degrees of freedom. ns - Not significant.

Figure 1. Matrix with frequency histograms (in the diagonal) and scatter plots (out of the diagonal) between plant height (PH), stem diameter (SD), number of nodes (NN), number of leaves (NL), leaf fresh matter (LFM), stem fresh matter (STFM), shoot fresh matter (SHFM = LFM + STFM), leaf dry matter (LDM), stem dry matter (STDM) and shoot dry matter (SHDM = LDM + STDM), measured at 147 days after sowing in 110 slender leaf rattlebox (*C. ochroleuca*) plants.

The linear association of PH with the traits LFM, STFM, SHFM, LDM, STDM and SHDM showed r values of low magnitude, that is, between 0.189 and 0.450, with mean of 0.352 (Table 3). Thus, it can be inferred that slender leaf rattlebox plants with greater height are associated with plants with more leaf, stem and shoot fresh and dry matter. The linear association of NN with the traits LFM, STFM, SHFM, LDM, STDM and SHDM was also of low magnitude ($0.243 \leq r \leq 0.359$, with mean = 0.323), showing that plants with more nodes are associated with plants with more leaf, stem and shoot fresh and dry matter.

Intermediate magnitude of association was observed in relation to SD and the traits LFM, STFM, SHFM, LDM, STDM and SHDM ($0.618 \leq r \leq 0.847$, with mean = 0.756), evidencing that plants with larger stem diameter are associated with plants with more leaf, stem and shoot fresh and dry matter. Furthermore, a linear association with greater magnitude was verified between NL and the traits LFM, STFM, SHFM, LDM, STDM and SHDM (0.762 ≤ r ≤ 0.895, with mean = 0.827), which reveals that plants with more leaves are associated with plants with more leaf, stem and shoot fresh and dry matter.

Leaf, stem and shoot fresh and dry matter (LFM, STFM, SHFM, LDM, STDM and SHDM) showed a higher degree of positive linear association (higher r values) with NL (mean r = 0.827) and SD (mean r = 0.756), when compared to PH (mean r = 0.352) and NN (mean r = 0.323) (Table 3), which suggests that NL and SD would be more strongly associated with leaf, stem and shoot fresh and dry matter of slender leaf rattlebox.

However, through correlation coefficients alone, it is not possible to infer which of the traits (PH, SD, NN and NL) has a direct effect on the traits LFM, STFM, SHFM, LDM, STDM and SHDM. Thus, path analysis is an appropriate procedure to infer about the true cause-effect relationships among traits (Cruz et al., 2012; Cruz et al., 2014).

The diagnosis of multicollinearity in Pearson's linear correlation coefficient (r) matrix, among the explanatory variables PH, SD, NN and NL, revealed a condition number of 7.165 (Table 4). Therefore, the matrix showed weak multicollinearity, according to the criteria of Montgomery and Peck (1982). With this, it can be inferred that the adverse effect of multicollinearity is overcome and that the path analyses of the main traits LFM, STFM, SHFM, LDM, STDM and SHDM of slender leaf rattlebox, as a function of the explanatory traits PH, SD, NN and NL, were performed under appropriate conditions (Montgomery and Peck, 1982; Cruz et al., 2012; Cruz et al., 2014).

PH showed a positive linear correlation (0.189 ≤ r ≤ 0.450) with the six traits (LFM, STFM, SHFM, LDM, STDM and SHDM). However, the direct effects of PH (|-0.006| ≤ direct effect ≤ 0.196) on these six traits were of low magnitude and, therefore, the association is explained by the highest indirect effects via SD (0.092 ≤ indirect effect ≤ 0.243) and via NL (0.077 ≤ indirect effect ≤ 0.134).

Table 4. Estimates of Pearson's correlation coefficients (r) and direct and indirect effects (path analysis) of the traits plant height (PH), stem diameter (SD), number of nodes (NN) and number of leaves (NL) on the traits leaf fresh matter (LFM), stem fresh matter (STFM), shoot fresh matter (SHFM = LFM + STFM), leaf dry matter (LDM), stem dry matter (STDM) and shoot dry matter (SHDM = LDM + STDM), measured at 147 days after sowing in 110 slender leaf rattlebox (*C. ochroleuca*) plants.

Effect	Main variable					
	LFM	STFM	SHFM	LDM	STDM	SHDM
Direct of PH on	-0.028	0.163	0.125	-0.006	0.196	0.166
Indirect of PH via SD	0.116	0.243	0.221	0.092	0.194	0.181
Indirect of PH via NN	-0.006	-0.033	-0.028	-0.030	-0.039	-0.038
Indirect of PH via NL	0.130	0.077	0.091	0.134	0.092	0.100
Pearson's correlation (r)	0.212*	0.450*	0.409*	0.189*	0.443*	0.409*
Direct of SD on	0.261	0.545	0.497	0.207	0.436	0.406
Indirect of SD via PH	-0.012	0.073	0.055	-0.003	0.087	0.074
Indirect of SD via NN	-0.003	-0.019	-0.016	-0.018	-0.023	-0.022
Indirect of SD via NL	0.418	0.248	0.294	0.431	0.296	0.324
Pearson's correlation (r)	0.663*	0.847*	0.830*	0.618*	0.796*	0.782*
Direct of NN on	-0.010	-0.056	-0.047	-0.052	-0.066	-0.065
Indirect of NN via PH	-0.016	0.096	0.073	-0.004	0.115	0.098
Indirect of NN via SD	0.090	0.189	0.172	0.072	0.151	0.141
Indirect of NN via NL	0.220	0.130	0.155	0.227	0.156	0.170
Pearson's correlation (r)	0.284*	0.359*	0.353*	0.243*	0.356*	0.344*
Direct of NL on	0.759	0.449	0.533	0.783	0.537	0.588
Indirect of NL via PH	-0.005	0.028	0.021	-0.001	0.033	0.028
Indirect of NL via SD	0.144	0.300	0.274	0.114	0.240	0.224
Indirect of NL via NN	-0.003	-0.016	-0.014	-0.015	-0.019	-0.019
Pearson's correlation (r)	0.895*	0.762*	0.814*	0.881*	0.792*	0.821*
Coefficient of determination	0.844	0.858	0.881	0.804	0.835	0.846
Residual variable	0.156	0.142	0.119	0.196	0.165	0.154
Condition number	7.165	7.165	7.165	7.165	7.165	7.165

* Significant at 5% probability of error by Student's t-test, with 108 degrees of freedom.

NN also showed a positive linear correlation (0.243 ≤ r ≤ 0.359) with the six traits (LFM, STFM, SHFM, LDM, STDM and SHDM). However, the direct effects of NN (|-0.010| ≤ direct effect ≤ |-0.066|) on these six traits were negligible and, therefore, the association, again, is explained by the highest indirect effects via SD (0.072 ≤ indirect effect ≤ 0.189) and via NL (0.130 ≤ indirect effect ≤ 0.227).

SD showed positive linear correlations with LFM (r = 0.663) and with LDM (r = 0.618) and the direct effects on LFM (0.261) and LDM (0.207) were of low magnitude, so the association was explained by the highest indirect effects of SD via NL on LFM (0.418) and LDM (0.431). However, there were greater association of SD with the traits STFM, SHFM, STDM and SHDM (0.782 ≤ r ≤ 0.847) and direct effects (0.406 ≤ direct effect ≤ 0.545) with the same sign and of greater magnitude than the indirect effects of SD via NL (0.248 ≤ direct effect ≤ 0.324), confirming cause-effect relationship among SD and the traits STFM, SHFM, STDM and SHDM (Table 4).

NL showed positive linear correlations with LFM (r = 0.895) and with LDM (r = 0.881) and the direct effects on LFM (0.759) and LDM (0.783) had the same sign and magnitude. Association of NL with the traits STFM, SHFM, STDM and SHDM (0.762 ≤ r ≤ 0.821) and direct effects (0.449 ≤ direct effect ≤ 0.588) with the same sign and of greater magnitude than the indirect effects of NL via SD (0.224 ≤ indirect effect ≤ 0.300) were also observed. Thus, a cause-effect relationship between NL and the traits LFM, STFM, SHFM, LDM, STDM and SHDM (Table 4) is confirmed.

The strong linear association among the traits LFM, STFM, SHFM, LDM, STDM and SHDM (0.810 ≤ r ≤ 0.997, with mean of 0.919) (Table 3) explains the similar results of the six path analyses (Table 4). Positive and high-magnitude association between shoot fresh and dry matter has also been found in forage turnip (r = 0.9671), white lupine (r = 0.9828) (Cargnelutti Filho et al., 2014) and jack bean (r = 0.960) (Cargnelutti Filho et al., 2018a).

In practice, based on the results of this study, it can be inferred that plants with larger stem diameter and with more leaves are associated with higher LFM, STFM, SHFM, LDM, STDM and SHDM. Therefore, stem diameter and number of leaves can be used for indirect selection, aiming at increasing fresh and dry matter. The fact that it is not necessary to destroy the plants to measure stem diameter and count the number of leaves is advantageous because it allows, if desired, to keep the plants until they produce seeds. For direct selection it would be necessary to destroy the plants for weighing to obtain LFM, STFM, SHFM, LDM, STDM and SHDM. Linear relations among traits and possibility of indirect selection have also been verified in forage turnip and white lupine (Cargnelutti Filho et al., 2014), showy rattlebox (Toebe et al., 2017), pigeon pea (Cargnelutti Filho et al., 2017) and jack bean (Cargnelutti Filho et al., 2018b).

4. Conclusions

For estimating the means of plant height, stem diameter, number of nodes, number of leaves, leaf fresh matter, stem fresh matter, shoot fresh matter, leaf dry matter, stem dry matter and shoot dry matter of slender leaf rattlebox (*Crotalaria ochroleuca*), with maximum error of 10% of the mean and confidence level of 95%, 67 plants are needed. In an experiment, to estimate the mean of each treatment with 10% precision, 67 plants per treatment should be evaluated. Number of leaves has a positive linear relation with leaf fresh and dry matter. Number of leaves and stem diameter have a positive linear relation with stem and shoot fresh and dry matter.

Authors' Contribution

Alberto Cargnelutti Filho contributed to the organization of the research, statistical analysis of data, interpretation of results, writing of the manuscript and final correction of the manuscript. Daniela Lixinski Silveira, Valéria Escaio Bubans, Bruno Fillipin Osmari, Felipe Manfio Somavilla, Vithória Morena Ortiz e Lucas Fillipin Osmari contributed equally in the execution of the experiment, data collection and final correction of the manuscript.

Acknowledgments

To the National Council for Scientific and Technological Development (CNPq - Processes 304652/2017-2; 146258/2019-3; 159611/2019-9), the Coordination for the Improvement of Higher Education Personnel (CAPES) and the Rio Grande do Sul State Research Support Foundation (FAPERGS) for granting scholarships to the authors. To scholarship-holding students and volunteers for their assistance in data collection.

Bibliographic References

Alvares, C.A., Stape, J.L., Sentelhas, P.C., Gonçalves, J.L.M., Sparovek, G. 2013. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, 22(6), 711-728. DOI: https://dx.doi.org/10.1127/0941-2948/2013/0507

Burin, C., Cargnelutti Filho, A., Toebe, M., Alves, B.M., Fick, A.L. 2014. Dimensionamento amostral para a estimação da média e da mediana de caracteres de tremoço branco (*Lupinus albus* L.). Comunicata Scientiae, 5(2), 205-212.

https://comunicatascientiae.com.br/comunicata/article/view/317 (acessado 04 de janeiro de 2022).

Bussab, W.O., Morettin, P.A. 2017. Estatística Básica. São Paulo: Saraiva.

Cargnelutti Filho, A., Alves, B.M., Toebe, M., Burin, C., Wartha, C.A. 2018a. Sample size to estimate the mean of traits in jack bean. Revista Brasileira de Ciências Agrárias, 13(1), e5505. DOI: http://dx.doi.org/10.5039/agraria.v13i1a5505

Cargnelutti Filho, A., Facco, G., Santos, G.O., Alves, B.M., Burin, C., Kleinpaul, J.A. 2017. Linear relations among pigeon pea traits. Comunicata Scientiae, 8(3), 493-502. DOI: http://dx.doi.org/10.14295/CS.v8i3.1420

Cargnelutti Filho, A., Toebe, M., Alves, B.M., Burin, C., Wartha, C.A. 2018b. Linear relations among traits in jack bean (*Canavalia ensiformis*). Bioagro, 30(2), 157-162. https://revistas.uclave.org/index.php/bioagro/article/view/1291 (acessado 04 de janeiro de 2022).

Cargnelutti Filho, A., Toebe, M., Burin, C., Alves, B.M., Facco, G., Casarotto, G. 2014. Relações lineares entre caracteres de nabo forrageiro e de tremoço branco. Ciência Rural, 44(1), 18-24. DOI: https://doi.org/10.1590/S0103-84782014000100004

Cruz C.D., Carneiro P.C.S., Regazzi A.J. 2014. Modelos biométricos aplicados ao melhoramento genético: volume 2. 3ª ed. Viçosa: UFV.

Cruz C.D., Regazzi A.J., Carneiro P.C.S. 2012. Modelos biométricos aplicados ao melhoramento genético: volume 1. 4ª ed. Viçosa: UFV.

Cruz, C.D. 2016. Genes Software - extended and integrated with the R, Matlab and Selegen. Acta Scientiarum Agronomy, v.38(4), 547-552. DOI: https://doi.org/10.4025/actasciagron.v38i4.32629

Facco, G., Cargnelutti Filho, A., Alves, B.M., Burin, C., Santos, G.O., Kleinpaul, J.A., Neu, I.M.M. 2016. Sample size for estimating average productive traits of pigeon pea. Ciência Rural, 46(4), 619-625. DOI: http://dx.doi.org/10.1590/0103-8478cr20150852

Facco, G., Cargnelutti Filho, A., Lúcio, A.D., Santos, G.O., Stefanello, R.B., Alves, B.M., Burin, C., Neu, I.M.M., Kleinpaul, J.A. 2015. Sample size for morphological traits of pigeonpea. Semina: Ciências Agrárias, 36(6), 4151-4164. DOI: http://dx.doi.org/10.5433/1679-0359.2015v36n6Supl2p4151

Ferreira D.F. 2009. Estatística básica, 2ª ed., Lavras: UFLA.

Montgomery D.C., Peck E.A. 1982. Introduction to linear regression analysis. New York: John Wiley & Sons.

Santos, H.G.; Jacomine, P.K.T.; Anjos, L.H.C.; Oliveira, V.A.; Lumbreras, J.F.; Coelho, M.R.; Almeida, J.A.; Araújo Filho, J.C.; Oliveira, J.B.; Cunha, T.J.F. 2018. Sistema Brasileiro de Classificação de Solos, 5ª ed., Brasília: Embrapa. https://www.embrapa.br/solos/sibcs (acessado 04 de janeiro de 2022).

Schabarum, D.E., Cargnelutti Filho, A., Bem, C.M., Facco, G., Kleinpaul, J.A., Wartha, C.A. 2018a. Sample sufficiency for mean estimation of productive traits of sunn hemp. Journal of Agricultural Science; 10(9), 209-216. DOI: https://doi.org/10.5539/jas.v10n9p209

Schabarum, D.E., Cargnelutti Filho, A., Lavezo, A., Follmann, D.N., Kleinpaul, J.A., Chaves, G.G., Pezzini, R.V. 2018b. Sample size for morphological traits of sunn hemp. Journal of Agricultural Science; 10(1), 152-161. DOI: https://doi.org/10.5539/jas.v10n1p152

Storck, L., Garcia, D.C. Lopes, S.J. Estefanel, V. 2016. Experimentação vegetal, 3ª ed., Santa Maria: UFSM.

Teodoro, P.E., Silva Junior, C.A., Ribeiro, L.P., Silva, F.A., Corrêa, C.C.G., Zanuncio, A.S., Torres, F.E. 2015. Sample dimension for estimation of biomass and yield of sunn (*Crotalaria juncea* L.) and showy rattlebox (*C. spectabilis* Roth.). Journal of Agronomy, 14(2), 98-101. DOI: http://dx.doi.org/10.3923/ja.2015.98.101

Toebe, M., Bandeira, C.T., Fortes, S.K.G., Carvalho, J.O., Tartaglia, F.L., Tambara, A.L., Melo, P.J. 2017. Dimensionamento amostral e associação linear entre caracteres de *Crotalaria spectabilis*. Bragantia, 76(1), 45-53. DOI: http://dx.doi.org/10.1590/1678-4499.653

Toebe, M., Machado, L.N., Tartaglia, F.L., Carvalho, J.O., Bandeira, C.T., Cargnelutti Filho, A. 2018. Sample size for estimating mean and coefficient of variation in species of crotalarias. Anais da Academia Brasileira de Ciências, 90(2), 1705-1715. DOI: http://dx.doi.org/10.1590/0001-3765201820170813