SEMINA
Ciências Agrárias

DOI: 10.5433/1679-0359.2023v44n1p203

Sampling sufficiency, correlation and path analysis in showy rattlebox

Suficiência amostral, correlação e análise de trilha em crotalária spectabilis

Alberto Cargnelutti Filho[1]*; Daniela Lixinski Silveira[2]; Felipe Manfio Somavilla[3]; Bruno Fillipin Osmari[3]; Vithória Morena Ortiz[3]

Highlights

Ten traits were evaluated in 310 plants of *Crotalaria spectabilis.*
Sampling sufficiency was determined and the linear relations were evaluated.
Information was generated for precise experimental planning.

Abstract

It is important to adequately size the number of plants that should be evaluated to allow precise inferences about the traits under evaluation. The study of the linear relations among traits provides important information, especially in the identification of traits for indirect selection. So, the objectives of this work were to determine the sample size (number of plants) to estimate the mean of *Crotalaria spectabilis* traits and investigate the relations among traits. Were randomly selected 200 and 110 plants of *C. spectabilis* in the experiments conducted, respectively, in 2019/2020 and 2020/2021. In these 310 plants, the following traits were evaluated: 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. The sample size was calculated to estimate the mean of these traits, based on Student's t-distribution, and the relations among traits were investigated through correlation and path analysis. To estimate the mean of these ten traits of *C. spectabilis,* with a maximum error of 10% of the mean and 95% confidence level, 64 plants are needed. In an experiment, to estimate the mean of each treatment with 10% precision, 64 plants per treatment must be evaluated. The number of leaves has a positive linear relation with the amount leaf, stem and shoot fresh and dry matter.

Key words: *Crotalaria spectabilis.* Linear relations. Number of plants. Sample sizing.

[1] Prof., Universidade Federal de Santa Maria, UFSM, Santa Maria, RS, Brazil. E-mail: alberto.cargnelutti.filho@gmail.com
[2] Student of the Doctoral Course of the Graduate Program in Agronomy, UFSM, Santa Maria, RS, Brazil. E-mail: danilisil@gmail.com
[3] Students of the Undergraduate Course in Agronomy, UFSM, Santa Maria, RS, Brazil. E-mail: felipe-somavilla@hotmail.com; brunoosmari11@gmail.com; vithoria.ortiz159@gmail.com
* Author for correspondence

Received: May 19, 2022 Approved: Feb. 28, 2023

Resumo

É importante dimensionar adequadamente o número de plantas que devem ser avaliadas para possibilitar inferências precisas sobre os caracteres em avaliação. O estudo das relações lineares entre caracteres fornece informações importantes, especialmente, na identificação de caracteres para seleção indireta. Assim, os objetivos deste trabalho foram determinar o tamanho de amostra (número de plantas) necessário para a estimação da média de caracteres de *Crotalaria spectabilis* e investigar as relações entre os caracteres. Foram selecionadas, aleatoriamente, 200 e 110 plantas de *C. spectabilis,* nos experimentos conduzidos, respectivamente, em 2019/2020 e 2020/2021. Nessas 310 plantas avaliaram-se os 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. Foi calculado o tamanho de amostra para a estimação da média desses caracteres, com base na distribuição t de *Student* e investigada a relação entre os caracteres por meio de análises de correlação e de trilha. Para a estimação da média, desses dez caracteres de *C. spectabilis,* com erro máximo de 10% da média e grau de confiança de 95%, são necessárias 64 plantas. Em um experimento, para a estimação da média de cada tratamento com 10% de precisão, devem ser avaliadas 64 plantas por tratamento. O número de folhas tem relação linear positiva com a quantidade de matérias fresca e seca de folhas, do caule e de parte aérea.

Palavras-chave: *Crotalaria spectabilis.* Dimensionamento amostral. Número de plantas. Relações lineares.

Introduction

Showy rattlebox (*Crotalaria spectabilis* Roth.) is an annual spring-summer legume. The species stands out for its efficiency in reducing the population of nematodes, especially those which cause galls (*Meloidogyne incognita*), cysts (*Heterodera glycine*) and root lesions (*Pratylenchus brachyurus*). It has high fresh matter production and excellent biological nitrogen fixation capacity. Its shoot biomass has low carbon:nitrogen ratio (18-20), which favors decomposition and nutrient cycling. The root system is branched and deep, favoring the absorption of nutrients leached in the soil profile, mainly potassium, calcium, magnesium and nitrate (Carvalho et al., 2022).

Due to its importance, experiments with this species are conducted in the field. In these experiments, the evaluation of all plants (individuals) of the usable area of the experimental unit (plot) is adequate to represent the trait under study. However, due to limitations of time, labor and financial resources, it is common to evaluate part of the plants (sample) of the plot, and the sample must be representative of the plants in the experimental unit (Storck et al., 2016). Thus, it is important to adequately size the number of plants that should be evaluated to allow precise inferences about the traits under evaluation.

Sample size is directly proportional to the variation of the data and to the desired confidence level in the estimate and inversely proportional to the estimation error allowed (Ferreira, 2009; Bussab & Morettin, 2017). The sampling sufficiency, for estimating the mean of traits, has been determined for traits of species of the Fabaceae family, such as:

Lupinus albus (Burin et al., 2014), *Dolichos lablab* (Teodoro et al., 2014), *Cajanus cajan* (Facco et al., 2015, 2016), *Crotalaria juncea* (Teodoro et al., 2015; Schabarum et al., 2018a,b; Toebe et al., 2018), *Crotalaria spectabilis* (Teodoro et al., 2015; Toebe et al., 2017, 2018), *Crotalaria breviflora, Crotalaria ochroleuca* (Toebe et al., 2018) and *Canavalia ensiformis* (Teodoro et al., 2014; Cargnelutti et al., 2018a). Variation in sample size between traits and species has been found in these studies.

The study of the interrelations among traits provides important information, especially in the identification of traits for indirect selection. Linear relations among a set of traits can be investigated through Pearson's linear correlation coefficient (r) and complementary statistical procedures, such as path analysis. 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 & 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 identify if there is linear association of cause and effect, enabling the identification of traits for indirect selection (Cruz et al., 2012, 2014). These statistical procedures have been used to study the linear association among traits of *Raphanus sativus* and *Lupinus albus* (Cargnelutti et al., 2014), *Crotalaria spectabilis* (Toebe et al., 2017), *Cajanus cajan* (Cargnelutti et al., 2017) and *Canavalia ensiformis* (Cargnelutti et al., 2018b).

The inclusion of *Crotalaria spectabilis* traits not contemplated in previous studies adds important information to support the planning of experiments with better precision. Thus, the objectives of this work were to determine the sample size (number of plants) required to estimate the mean of traits of showy rattlebox (*Crotalaria spectabilis* Roth.) and investigate the relations among traits.

Material and Methods

Two uniformity trials (blank experiments) were conducted with showy rattlebox (*Crotalaria spectabilis* Roth.), in Santa Maria, State of Rio Grande do Sul, Brazil (29º42'S latitude, 53º49'W longitude and 95 m altitude), in an area of 8 m × 20 m (160 m²). In this place, the climate is humid subtropical - Cfa (Alvares et al., 2013) and the soil is *Argissolo Vermelho distrófico arênico* (Ultisol) (Santos et al., 2018).

The first uniformity trial was carried out in the agricultural year of 2019/2020. In this trial, on November 13, 2019, 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}. The second uniformity trial was carried out in the agricultural year of 2020/2021. In this trial, 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 performed broadcast, with 22.50 kg of seeds ha^{-1}.

Plants were collected when they were at the end of the vegetative stage and beginning of the reproductive stage. In the first trial, at 105 days after sowing (DAS), 200 plants were randomly selected. In the second trial, at 85 DAS, 110 plants were randomly selected.

Each plant was evaluated for the following traits: plant height (PH, in cm), 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 (PH, SD, NN, NL, LFM, STFM, SHFM, LDM, STDM and SHDM), measures of central tendency, dispersion, skewness, kurtosis and p-value of the Kolmogorov-Smirnov normality test were calculated.

For each trait, based on the number of plants sampled, that is, 200 plants for the first trial and 110 plants for the second trial, sample size (n) was calculated for the estimation errors (semi-amplitudes of the confidence interval) fixed at 1%, 2%, ..., 20% of the mean (m), that is, 0.01×m (higher precision), 0.02×m, ... 0.20×m (lower precision), with a 95% confidence level (1-α) by the expression $n = \left(\frac{t_{\alpha/2}\, s}{estimation\ error}\right)^2$ (Ferreira, 2009; Bussab & Morettin, 2017), in which $t_{\alpha/2}$ 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% significance and n-1 degrees of freedom, and s is the estimate of the standard deviation.

To investigate the relations among the traits PH, SD, NN, NL, LFM, STFM, SHFM, LDM, STDM and SHDM, Pearson's linear correlation coefficient matrix (r) was determined and Student's t-test was used to check the significance of r at 5%. In the correlation matrix among the traits PH, SD, NN and NL, the diagnosis of multicollinearity was made (Cruz et al., 2014) and interpretation was performed according to the condition number (CN), which is the ratio between the highest and lowest eigenvalue of the correlation matrix. Multicollinearity was classified according to Montgomery & Peck (1982) as: weak when CN ≤ 100, moderate to severe when 100 < CN < 1000 and severe when CN ≥ 1000.

After that, path analysis of the main variables (LFM, STFM, SHFM, LDM, STDM and SHDM) as a function of the explanatory variables (PH, SD, NN and NL) was performed by following the methodology described in Cruz et al. (2012, 2014), in a total of six path analyses, for each uniformity trial. For example, in the first trial, the first path analysis comprised LFM as a function of PH, SD, NN and NL, the second path analysis comprised STFM as a function of PH, SD, NN and NL, and so on, up to the sixth path analysis, which comprised SHDM as a function of PH, SD, NN and NL. Statistical analyses were carried out with the applications Microsoft Office Excel® and Genes (Cruz, 2016).

Results and Discussion

For the data of 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 between 0.000 and 0.802, with mean of 0.199 in the agricultural year of 2019/2020 and between 0.001 and 0.877, with mean of 0.322 in the agricultural year of 2020/2021 (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 2.5%, it can be inferred that the assumption of normality was met for the traits PH, NN, NL, LFM, STFM, SHFM, LDM, STDM and SHDM. For this test, the SD showed low adherence to the normal distribution (p-value ≤ 0.001). However, the kurtosis not different from three and the proximity of the mean to the median are indicative of a slight distance from the normal distribution curve. Thus, this data set is considered to be adequate for the studies of sample sizing based on Student's t-distribution and linear relations through correlation and path analyses.

Based on the mean of the traits, there was greater growth and development of plants in the agricultural year of 2019/2020, despite the lower NL (Table 1). The highest means of the traits in the agricultural year of 2019/2020 may be associated with the longer permanence of the plants in the field. In the agricultural year of 2019/2020, the plants were collected at 105 days after sowing (DAS), while in the agricultural year of 2020/2021 they were collected at 85 DAS. Based on the dispersion measures, there was variation for all traits between the plants sampled within each agricultural year. The variation between plants within the agricultural year and between agricultural years 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.

In both agricultural years, it was observed that the coefficients of variation (CV) for the traits NL, LFM, STFM, SHFM, LDM, STDM and SHDM (29.97% ≤ CV ≤ 40.27%, with mean of 35.14%) were relatively higher than those of PH, SD and NN (6.06% ≤ CV ≤ 20.74%, with mean of 11.84%) (Table 1). Thus, for the same precision, a larger sample size is expected to estimate the means 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 143 plants for PH in the agricultural year of 2019/2020 to 6307 plants for LFM in the agricultural year of 2019/2020 (Table 2). In Microsoft Office Excel®, these sizes are obtained, respectively, by the following expressions: =ARREDONDAR.PARA.CIMA((((INVT(0.05;199)*10.01)/(0.01*165.10))^2);0) = 143 plants and =ARREDONDAR.PARA.CIMA((((INVT(0.05;199)*27.6477)/(0.01*68.6521))^2);0)= 6307 plants.

Table 1

Minimum, median, mean, maximum, amplitude, variance, standard deviation, standard error, coefficient of variation (CV), skewness, kurtosis and p-value of the Kolmogorov-Smirnov normality test of traits[(1)] of showy rattlebox (*Crotalaria spectabilis* Roth.) plants, in two agricultural years

Statistics	PH	SD	NN	NL	LFM	STFM	SHFM	LDM	STDM	SHDM
Agricultural year of 2019/2020 (n = 200 plants sampled)										
Minimum	139.00	0.80	37.00	16.00	20.27	50.34	70.61	3.25	7.44	10.69
Median	165.00	1.20	47.00	57.00	63.66	151.25	213.75	11.92	28.79	41.80
Mean	165.10	1.21	46.35	60.46	68.65	160.65	229.31	12.65	30.08	42.73
Maximum	196.00	1.90	57.00	144.00	175.70	349.19	522.95	30.10	63.43	92.35
Amplitude	57.00	1.10	20.00	128.00	155.43	298.85	452.34	26.85	55.99	81.66
Variance	100.13	0.06	16.02	590.23	764.39	3291.38	6927.81	22.63	112.20	227.54
Standard deviation	10.01	0.25	4.00	24.29	27.65	57.37	83.23	4.76	10.59	15.08
Standard error	0.71	0.02	0.28	1.72	1.95	4.06	5.89	0.34	0.75	1.07
CV (%)	6.06	20.74	8.63	40.18	40.27	35.71	36.30	37.59	35.22	35.30
Skewness [(2)]	0.11ns	0.72*	-0.12ns	0.95*	1.27*	0.73*	0.84*	1.06*	0.65*	0.72*
Kurtosis + 3 [(3)]	3.41ns	3.17ns	2.90ns	3.83*	5.09*	3.34ns	3.62ns	4.36*	3.10ns	3.25ns
p-value	0.802	0.000	0.079	0.040	0.025	0.163	0.130	0.129	0.287	0.337
Agricultural year of 2020/2021 (n = 110 plants sampled)										
Minimum	56.00	0.60	21.00	32.00	21.34	44.03	78.56	4.73	8.44	14.77
Median	74.50	1.00	27.00	70.50	54.70	97.36	155.57	11.47	18.36	29.72
Mean	74.55	1.03	26.87	76.53	59.05	102.93	161.98	12.13	19.23	31.37
Maximum	94.00	1.40	32.00	159.00	121.15	196.66	299.88	23.40	34.33	56.68
Amplitude	38.00	0.80	11.00	127.00	99.81	152.63	221.32	18.67	25.89	41.91
Variance	77.70	0.02	5.52	905.76	440.65	951.53	2453.06	17.51	35.33	92.27
Standard deviation	8.81	0.15	2.35	30.10	20.99	30.85	49.53	4.18	5.94	9.61
Standard error	0.84	0.01	0.22	2.87	2.00	2.94	4.72	0.40	0.57	0.92
CV (%)	11.82	15.02	8.75	39.33	35.55	29.97	30.58	34.49	30.90	30.62
Skewness [(2)]	0.05ns	-0.16ns	-0.11ns	0.71*	0.74*	0.64*	0.62*	0.60*	0.48*	0.47*
Kurtosis + 3 [(3)]	2.53ns	2.90ns	2.56ns	2.74ns	2.95ns	3.07ns	2.73ns	2.64ns	2.66ns	2.49ns
p-value	0.877	0.001	0.199	0.202	0.110	0.365	0.239	0.324	0.508	0.400

[(1)] PH - plant height, in cm; 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 the Student's t-test, at 5% significance level. ns Not significant. [(3)] * Kurtosis differs from three, through the Student's t-test, at 5% significance level. ns Not significant.

Therefore, in relation to plant height, it can be inferred, with 95% confidence, that the mean confidence interval (μ), obtained with 143 plants, is μ±0.01m, i.e., μ±1.651 cm, because the mean height of the 200 plants sampled was 165.10 cm (Table 1). In the other extreme, the precision of μ±0.01m is obtained with 6307 plants for leaf fresh matter and, in this situation, the value would have been μ±0.686 grams, because the mean LFM of the 200 plants sampled was 68.65 grams. These results reveal that, for the same precision, the sample sizes vary among the traits, as verified in *Lupinus albus* (Burin et al., 2014), *Dolichos lablab* (Teodoro et al., 2014), *Cajanus cajan* (Facco et al., 2015, 2016), *Crotalaria juncea* (Teodoro et al., 2015; Schabarum et al., 2018a,b; Toebe et al., 2018), *Crotalaria spectabilis* (Teodoro et al., 2015; Toebe et al., 2017, 2018), *Crotalaria breviflora, Crotalaria ochroleuca* (Toebe et al., 2018) and *Canavalia ensiformis* (Teodoro et al., 2014; Cargnelutti et al., 2018a).

Table 2

Sample size (number of plants) for estimating the means of traits(1) of showy rattlebox (*Crotalaria spectabilis* Roth.), in two agricultural years, for estimation errors (semi-amplitudes of the confidence interval) equal to 1%, 2%, ..., 20% of the mean (m), that is, 0.01×m (higher precision), 0.02×m, ... 0.20×m (lower precision), with 95% confidence level (1-α)

Error	PH	SD	NN	NL	LFM	STFM	SHFM	LDM	STDM	SHDM
Agricultural year of 2019/2020 (n = 200 plants sampled)										
1%	143	1673	290	6279	6307	4959	5124	5497	4823	4846
2%	36	419	73	1570	1577	1240	1281	1375	1206	1212
3%	16	186	33	698	701	551	570	611	536	539
4%	9	105	19	393	395	310	321	344	302	303
5%	6	67	12	252	253	199	205	220	193	194
6%	4	47	9	175	176	138	143	153	134	135
7%	3	35	6	129	129	102	105	113	99	99
8%	3	27	5	99	99	78	81	86	76	76
9%	2	21	4	78	78	62	64	68	60	60
10%	2	17	3	63	64	50	52	55	49	49
11%	2	14	3	52	53	41	43	46	40	41
12%	1	12	3	44	44	35	36	39	34	34
13%	1	10	2	38	38	30	31	33	29	29
14%	1	9	2	33	33	26	27	29	25	25
15%	1	8	2	28	29	23	23	25	22	22
16%	1	7	2	25	25	20	21	22	19	19
17%	1	6	2	22	22	18	18	20	17	17
18%	1	6	1	20	20	16	16	17	15	15
19%	1	5	1	18	18	14	15	16	14	14
20%	1	5	1	16	16	13	13	14	13	13

continue...

continuation...

Agricultural year of 2020/2021 (n = 110 plants sampled)										
1%	550	887	301	6076	4965	3528	3673	4674	3752	3684
2%	138	222	76	1519	1242	882	919	1169	938	921
3%	62	99	34	676	552	392	409	520	417	410
4%	35	56	19	380	311	221	230	293	235	231
5%	22	36	13	244	199	142	147	187	151	148
6%	16	25	9	169	138	98	103	130	105	103
7%	12	19	7	124	102	72	75	96	77	76
8%	9	14	5	95	78	56	58	74	59	58
9%	7	11	4	76	62	44	46	58	47	46
10%	6	9	4	61	50	36	37	47	38	37
11%	5	8	3	51	42	30	31	39	32	31
12%	4	7	3	43	35	25	26	33	27	26
13%	4	6	2	36	30	21	22	28	23	22
14%	3	5	2	31	26	18	19	24	20	19
15%	3	4	2	28	23	16	17	21	17	17
16%	3	4	2	24	20	14	15	19	15	15
17%	2	4	2	22	18	13	13	17	13	13
18%	2	3	1	19	16	11	12	15	12	12
19%	2	3	1	17	14	10	11	13	11	11
20%	2	3	1	16	13	9	10	12	10	10

(1) PH - plant height, in cm; 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 plant^{-1} g; 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}.

With this estimation error of 1%, a smaller sample size was observed for the traits PH, SD and NN (143 ≤ n ≤ 1673, with mean of 641 plants) when compared to the sample size required for the traits NL, LFM, STFM, SHFM, LDM, STDM and SHDM (3528 ≤ n ≤ 6307, with mean of 4871 plants) (Table 2). These results are due to the higher coefficients of variation of the traits NL, LFM, STFM, SHFM, LDM, STDM and SHDM compared to those of the traits PH, SD and NN (Table 1).

Based on the results of the agricultural year of 2019/2020, opting for allowing estimation error of 20%, that is, 0.20×m, and confidence level of 95%, the number of plants to be sampled decreases to 1, 5, 1, 16, 16, 13, 13, 14, 13 and 13, 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 (≤ 16) would be easily evaluated in an experiment. However, it would lead to less precision in estimating the mean of the traits, i.e. μ±33.02 cm, μ±0.24

cm, µ±9.27 nodes, µ±12.09 leaves, µ±13.73 grams, µ±32.13 grams, µ±45.86 grams, µ±2.53 grams, µ±6.02 grams and µ±8.55 grams, respectively, for the traits PH, SD, NN, NL, LFM, STFM, SHFM, LDM, STDM and SHDM. This same reasoning can be applied to the results of the agricultural year of 2020/2021.

The results of this work allow the researcher to choose the sample size, 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, 64 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, 64 plants per treatment should be evaluated. If the experiment is planned with four replicates per treatment, 16 plants per replicate (64/4 = 16) would be sampled, that is, 16 plants per plot. Furthermore, if ten treatments were evaluated in the experiment, 640 plants (64 per treatment) would be sampled. It is worth pointing out that, for the traits PH, SD, NN, NL, individual evaluations of the 16 plants in the plot are required, while for the traits LFM, STFM, SHFM, LDM, STDM and SHDM, the weighing of the 16 plants in the plot can be performed together, which can facilitate the evaluation.

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 (0.858 ≤ r ≤ 0.942, with mean = 0.909) and decreasing association with the other traits in the following order: SD (0.693 ≤ r ≤ 0.802, with mean = 0.765) and NN (0.452 ≤ r ≤ 0.611, with mean = 0.530), which suggests that the number of leaves would be more strongly associated with the leaf, stem and shoot fresh and dry matter of *Crotalaria spectabilis.* On the other hand, there was a weak negative linear association or lack of linear association with PH (Table 3). Path analysis makes it possible to decompose the r into direct and indirect effects, making it possible to infer which explanatory trait (PH, SD, NN and NL) has a cause-effect relationship with the traits LFM, STFM, SHFM, LDM, STDM and SHDM (Cruz et al., 2012, 2014).

The low values of the condition number (CN ≤ 12.4) indicate that Pearson's linear correlation coefficient matrix (r) between the explanatory variables PH, SD, NN and NL showed weak multicollinearity (Cruz et al., 2014; Cruz, 2016) (Table 4). Therefore, the adverse effect of multicollinearity is overcome and the path analyses of the main traits LFM, STFM, SHFM, LDM, STDM and SHDM of *Crotalaria spectabilis,* as a function of the explanatory traits PH, SD, NN and NL, in the two agricultural years, were performed under appropriate conditions (Cruz et al., 2012, 2014; Cruz, 2016).

Table 3

Estimates of Pearson's linear correlation coefficients among the traits[1] measured in showy rattlebox (*Crotalaria spectabilis* Roth.) plants, in two agricultural years (agricultural year of 2019/2020 above the diagonal and agricultural year of 2020/2021 below the diagonal)

	PH	SD	NN	NL	LFM	STFM	SHFM	LDM	STDM	SHDM
PH		-0.174*	0.175*	-0.299*	-0.254*	-0.105ns	-0.157*	-0.249*	-0.129ns	-0.169*
SD	0.219*		0.545*	0.771*	0.763*	0.782*	0.793*	0.764*	0.799*	0.802*
NN	0.535*	0.439*		0.513*	0.500*	0.611*	0.587*	0.508*	0.604*	0.585*
NL	-0.090ns	0.698*	0.443*		0.942*	0.889*	0.926*	0.934*	0.888*	0.918*
LFM	-0.001ns	0.720*	0.452*	0.905*		0.905*	0.956*	0.977*	0.893*	0.935*
STFM	0.059ns	0.766*	0.504*	0.880*	0.819*		0.990*	0.909*	0.980*	0.975*
SHFM	0.037ns	0.782*	0.506*	0.932*	0.934*	0.970*		0.951*	0.972*	0.982*
LDM	0.026ns	0.693*	0.469*	0.906*	0.979*	0.807*	0.918*		0.920*	0.961*
STDM	0.107ns	0.752*	0.512*	0.858*	0.779*	0.970*	0.934*	0.793*		0.992*
SHDM	0.077ns	0.767*	0.521*	0.926*	0.909*	0.952*	0.978*	0.926*	0.964*	

[1] PH - plant height, in cm; 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 plant^{-1} g; 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% by Student's t-test, with 198 and 108 degrees of freedom, respectively, for the agricultural years of 2019/2020 and 2020/2021. ns Not significant.

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) in showy rattlebox (*Crotalaria spectabilis* Roth.) plants in two agricultural years

Effect	Main variable					
	LFM	STFM	SHFM	LDM	STDM	SHDM
	Agricultural year of 2019/2020 (n = 200 plants sampled)					
Direct of PH on	0.028	0.123	0.094	0.024	0.096	0.075
Indirect of PH via SD	-0.016	-0.031	-0.026	-0.018	-0.039	-0.033
Indirect of PH via NN	-0.001	0.021	0.014	0.001	0.020	0.014
Indirect of PH via NL	-0.265	-0.218	-0.238	-0.257	-0.205	-0.225
Pearson's correlation (r)	-0.254*	-0.105ns	-0.157*	-0.249*	-0.129ns	-0.169*
Direct of SD on	0.089	0.177	0.151	0.102	0.225	0.190
Indirect of SD via PH	-0.005	-0.021	-0.016	-0.004	-0.017	-0.013
Indirect of SD via NN	-0.004	0.065	0.044	0.004	0.062	0.045
Indirect of SD via NL	0.682	0.562	0.614	0.661	0.528	0.579

continue...

continuation...

Pearson's correlation (r)	0.763*	0.782*	0.793*	0.764*	0.799*	0.802*
Direct of NN on	-0.008	0.119	0.080	0.008	0.113	0.082
Indirect of NN via PH	0.005	0.021	0.016	0.004	0.017	0.013
Indirect of NN via SD	0.049	0.096	0.083	0.056	0.123	0.104
Indirect of NN via NL	0.454	0.374	0.408	0.440	0.351	0.385
Pearson's correlation (r)	0.500*	0.611*	0.587*	0.508*	0.604*	0.585*
Direct of NL on	0.885	0.729	0.796	0.858	0.685	0.752
Indirect of NL via PH	-0.008	-0.037	-0.028	-0.007	-0.029	-0.022
Indirect of NL via SD	0.069	0.136	0.117	0.079	0.174	0.147
Indirect of NL via NN	-0.004	0.061	0.041	0.004	0.058	0.042
Pearson's correlation (r)	0.942*	0.889*	0.926*	0.934*	0.888*	0.918*
Coefficient of determination	0.891	0.846	0.890	0.877	0.845	0.878
Residual variable	0.109	0.154	0.110	0.123	0.155	0.122
Condition number	10.4	10.4	10.4	10.4	10.4	10.4
	Agricultural year of 2020/2021 (n = 110 plants sampled)					
Direct of PH on	0.026	0.012	0.018	0.077	0.078	0.082
Indirect of PH via SD	0.033	0.058	0.050	0.016	0.052	0.039
Indirect of PH via NN	0.011	0.048	0.035	0.009	0.037	0.027
Indirect of PH via NL	-0.071	-0.059	-0.067	-0.077	-0.060	-0.070
Pearson's correlation (r)	-0.001ns	0.059ns	0.037ns	0.026ns	0.107ns	0.077ns
Direct of SD on	0.153	0.266	0.230	0.073	0.238	0.179
Indirect of SD via PH	0.006	0.003	0.004	0.017	0.017	0.018
Indirect of SD via NN	0.009	0.040	0.028	0.008	0.030	0.022
Indirect of SD via NL	0.553	0.458	0.519	0.596	0.467	0.548
Pearson's correlation (r)	0.720*	0.766*	0.782*	0.693*	0.752*	0.767*
Direct of NN on	0.020	0.090	0.065	0.017	0.069	0.050
Indirect of NN via PH	0.014	0.006	0.010	0.041	0.042	0.044
Indirect of NN via SD	0.067	0.117	0.101	0.032	0.105	0.079
Indirect of NN via NL	0.351	0.291	0.330	0.379	0.296	0.348
Pearson's correlation (r)	0.452*	0.504*	0.506*	0.469*	0.512*	0.521*
Direct of NL on	0.792	0.656	0.744	0.854	0.668	0.786
Indirect of NL via PH	-0.002	-0.001	-0.002	-0.007	-0.007	-0.007
Indirect of NL via SD	0.107	0.186	0.161	0.051	0.166	0.125
Indirect of NL via NN	0.009	0.040	0.029	0.008	0.031	0.022
Pearson's correlation (r)	0.905*	0.880*	0.932*	0.906*	0.858*	0.926*
Coefficient of determination	0.836	0.827	0.907	0.834	0.796	0.897
Residual variable	0.164	0.173	0.093	0.166	0.204	0.103
Condition number	12.4	12.4	12.4	12.4	12.4	12.4

* Significant at 5% by Student's t-test, with 198 and 108 degrees of freedom, respectively, for the agricultural years of 2019/2020 and 2020/2021. ns Not significant.

The strong linear association between the traits LFM, STFM, SHFM, LDM, STDM and SHDM (0.779 ≤ r ≤ 0.992, with mean of 0.931) (Table 3) explains the similar results of the twelve path analyses (Table 4). A positive and high-magnitude association between shoot fresh and dry matter has also been observed in *Raphanus sativus* (r = 0.9671), *Lupinus albus* (r = 0.9828) (Cargnelutti et al., 2014) and *Canavalia ensiformis* (r = 0.960) (Cargnelutti et al., 2018b).

The direct effects of low magnitude of PH (0.012 ≤ direct effect ≤ 0.123) on the traits LFM, STFM, SHFM, LDM, STDM and SHDM confirm the absence of linear association of PH with these traits. SD showed a positive linear correlation (0.693 ≤ r ≤ 0.802) with the traits LFM, STFM, SHFM, LDM, STDM and SHDM and direct effects (0.073 ≤ direct effect ≤ 0.266) of low magnitude. Therefore, the association is explained by the greatest indirect effects via NL (0.458 ≤ indirect effect ≤ 0.682). NN also showed a positive linear correlation (0.452 ≤ r ≤ 0.611) with the traits LFM, STFM, SHFM, LDM, STDM and SHDM. However, the direct effects of NN (|-0.008| ≤ direct effect ≤ 0.119) on these six traits were negligible, and the association was explained again by the greatest indirect effects via NL (0.291 ≤ indirect effect ≤ 0.454) (Table 4).

NL showed a positive linear correlation (0.858 ≤ r ≤ 0.942) and direct effects (0.656 ≤ direct effect ≤ 0.885), with the same sign and similar magnitude, with the traits LFM, STFM, SHFM, LDM, STDM and SHDM. Thus, a cause-effect relationship between NL and the traits LFM, STFM, SHFM, LDM, STDM and SHDM (Table 4) is confirmed. Therefore, plants with more leaves have larger amounts of leaf, stem and shoot fresh and dry matter. The fact that it is not necessary to destroy the plants to count the number of leaves is advantageous, because it makes it possible to select plants aiming to increase the fresh and dry matter, keeping them until the production of seeds. For direct selection, it would be necessary to destroy the plants for weighing to obtain LFM, STFM, SHFM, LDM, STDM and SHDM. In *Raphanus sativus* and *Lupinus albus* (Cargnelutti et al., 2014), *Crotalaria spectabilis* (Toebe et al., 2017), *Cajanus cajan* (Cargnelutti et al., 2017) and *Canavalia ensiformis* (Cargnelutti et al., 2018b), cause-effect relationships among several traits and the possibility of indirect selection have also been found.

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 *Crotalaria spectabilis,* with maximum error of 10% of the mean and confidence level of 95%, 64 plants are required. In an experiment, to estimate the mean of each treatment with 10% precision, 64 plants per treatment should be evaluated. The number of leaves has a positive linear relation with leaf, stem and shoot fresh and dry matter.

Acknowledgments

To the National Council for Scientific and Technological Development (CNPq - Processes 304652/2017-2; 146258/2019-3), the Coordination for the Improvement of Higher Education Personnel (CAPES) and the Rio Grande do Sul State Research

Support Foundation (FAPERGS) for granting the scholarships. To scholarship-holding students and volunteers for their assistance in data collection.

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: 10.1127/0941-2948/2013/0507

Burin, C., Cargnelutti, A., Fº., 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. doi: 10.14295/cs.v5i2.317

Bussab, W. O., & Morettin, P. A. (2017). *Estatística básica.* Saraiva.

Cargnelutti, A., Fº., 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: 10.5039/agraria.v13i1a5505

Cargnelutti, A., Fº., 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: 10.14295/CS.v8i3.1420

Cargnelutti, A., Fº., 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.

Cargnelutti, A., Fº., 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: 10.1590/S0103-84782014000100004

Carvalho, M. L., Vanolli, B. S., Schiebelbein, B. E., Borba, D. A., Luz, F. B., Cardoso, G. M., Bortolo, L. S., Marostica, M. E. M., & Souza, V. S. (2022). *Guia prático de plantas de cobertura: aspectos filotécnicos e impactos sobre a saúde do solo.* ESALQ-USP.

Cruz, C. D., Carneiro, P. C. S., & Regazzi, A. J. (2014). *Modelos biométricos aplicados ao melhoramento genético* (vol. 2, 3a ed.). UFV.

Cruz, C. D., Regazzi, A. J., & Carneiro, P. C. S. (2012). *Modelos biométricos aplicados ao melhoramento genético* (vol. 1, 4a ed.). UFV.

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

Facco, G., Cargnelutti, A., Fº., 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: 10.1590/0103-8478cr20150852

Facco, G., Cargnelutti, A., Fº., 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: 10.5433/1679-0359.2015v36n6Supl2p4151

Ferreira, D. F. (2009). *Estatística básica* (2a ed.). UFLA.

Montgomery, D. C., & Peck, E. A. (1982). *Introduction to linear regression analysis.* 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, J. C., Fº., Oliveira, J. B., & Cunha, T. J. F. (2018). *Sistema brasileiro de classificação de solos* (5a ed.). EMBRAPA.

Schabarum, D. E., Cargnelutti, A., Fº., 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: 10.5539/jas.v10n9p209

Schabarum, D. E., Cargnelutti, A., Fº., 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: 10.5539/jas.v10n1p152

Storck, L., Garcia, D. C., Lopes, S. J., & Estefanel, V. (2016). *Experimentação vegetal* (3a ed.). UFSM.

Teodoro, P. E., Ribeiro, L. P., Silva, F. A., Corrêa, C. C. G., Luz, R. A. A., Jr., & Torres, F. E. (2014). Sample size to estimate biomass and productivity of *Canavalia ensiformis* and *Dolichos lablab. Revista de Ciências Agrárias, 37*(3), 348-353. doi: 10.19084/rca.16834

Teodoro, P. E., Silva, C. A., Jr., 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: 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: 10.1590/1678-4499.653

Toebe, M., Machado, L. N., Tartaglia, F. L., Carvalho, J. O., Bandeira, C. T., & Cargnelutti, A., Fº. (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: 10.1590/0001-3765201820170813