PAB
Pesquisa Agropecuária Brasileira

Statistics/ Original Article

ISSN 1678-3921

Journal homepage: www.embrapa.br/pab

For manuscript submission and journal contents, access: www.scielo.br/pab

Sample size for the estimation of Pearson's linear correlation in crotalaria species

Abstract – The objective of this work was to determine the necessary sample size to estimate Pearson's linear correlation coefficients of four species of crotalaria at precision levels. The experiment was carried out with *Crotalaria juncea*, *Crotalaria spectabilis*, *Crotalaria breviflora*, and *Crotalaria ochroleuca*, during the 2014/2015 crop year. Eight crotalaria traits were evaluated in 1,000 randomly collected pods per species. For each species, the correlation coefficients were estimated for the 28 pairs of traits, and the sample size necessary to estimate the correlation coefficients was determined at four precision levels [0.10, 0.20, 0.30, and 0.40 amplitudes of the 95% ($CI_{95\%}$) confidence interval] by resampling with replacement. The sample size varies between crotalaria species and, especially, between pairs of traits, as a function of the magnitude of the correlation coefficient. At a certain precision level, the smallest sample size is required to estimate the correlation coefficients between highly correlated traits and vice-versa. To estimate the correlation coefficients with $CI_{95\%}$ of 0.20, 10 to 440 pods are required, depending on the species, pairs of traits, and magnitude of the correlation coefficient.

Index terms: *Crotalaria*, linear relationships, resampling, sample precision.

Marcos Toebe[1] ✉, Letícia Nunes Machado[2], Francieli de Lima Tartaglia[3], Juliana Oliveira de Carvalho[4], Cirineu Tolfo Bandeira[3] and Alberto Cargnelutti Filho[3]

[1] Universidade Federal de Santa Maria, Departamento de Ciências Agronômicas e Ambientais, Linha 7 de Setembro, s/nº, BR-386, Km 40, CEP 98400-000 Frederico Westphalen, RS, Brazil. E-mail: m.toebe@gmail.com

[2] Universidade Federal do Pampa, Rua Luiz Joaquim de Sá Britto, s/nº, Promorar, CEP 97650-000 Itaqui, RS, Brazil. E-mail: leticia.nunes.machado@gmail.com

[3] Universidade Federal de Santa Maria, Departamento de Fitotecnia, Avenida Roraima, nº 1.000, Camobi, CEP 97105-900 Santa Maria, RS, Brazil. E-mail: francielitartaglia@hotmail.com, cirineu_bandeira@hotmail.com, alberto.cargnelutti.filho@gmail.com

[4] Universidade Federal do Rio Grande do Sul, Departamento de Biociências, Avenida Bento Gonçalves, nº 9.500, CEP 91501-970 Porto Alegre, RS, Brazil. E-mail: ju.olic@gmail.com

✉ Corresponding author

Received
September 5, 2018

Accepted
August 5, 2019

How to cite
TOEBE, M.; MACHADO, L.N.; TARTAGLIA, F. de L.; CARVALHO, J.O. de; BANDEIRA, C.T.; CARGNELUTTI FILHO, A. Sample size for the estimation of Pearson's linear correlation in crotalaria species. **Pesquisa Agropecuária Brasileira**, v.54, e01027, 2019. DOI: https://doi.org/10.1590/S1678-3921.pab2019.v54.01027.

Tamanho de amostra para estimar coeficientes de correlação linear de Pearson em espécies de crotalária

Resumo – O objetivo deste trabalho foi determinar o tamanho de amostra necessário para estimar os coeficientes de correlação linear de Pearson em quatro espécies de crotalária, em níveis de precisão. O experimento foi realizado com *Crotalaria juncea*, *Crotalaria spectabilis*, *Crotalaria breviflora* e *Crotalaria ochroleuca*, no ano agrícola 2014/2015. Oito características da crotalária foram avaliadas em 1.000 vagens coletadas aleatoriamente por espécie. Para cada espécie, estimaram-se os coeficientes de correlação para os 28 pares de características e determinou-se o tamanho de amostra necessário para a estimação dos coeficientes de correlação, em quatro níveis de precisão [amplitudes do intervalo de confiança de 95% ($CI_{95\%}$) de 0,10, 0,20, 0,30 e 0,40] por reamostragem com reposição. O tamanho de amostra varia entre as espécies de crotalária e, principalmente, entre os pares de características, em função da magnitude do coeficiente de correlação. Em determinado nível de precisão, o menor tamanho de amostra é necessário para a estimação de coeficientes de correlação de alta magnitude e vice-versa. Para estimar coeficientes de correlação com $CI_{95\%}$ de 0,20, são necessárias de 10 a 440 vagens, a depender da espécie, dos pares de características e da magnitude do coeficiente de correlação.

Termos para indexação: *Crotalaria*, relações lineares, reamostragem, precisão amostral.

Introduction

Crotalaria species, such as *C. juncea*, are used as cover plants in crop rotation systems with high production of fresh matter (Chaudhary, 2016) and nitrogen supply to the subsequent crops, positively influencing plant growth and productivity (Diniz et al., 2017; Elsaid & Silva, 2017). Other species of crotalaria, such as *C. spectabilis*, *C. breviflora*, and *C. ochroleuca*, can reduce the incidence or pest infestation, diseases and nematodes (Deberdt et al., 2015; Braz et al., 2016; Reigada et al., 2016).

Although crotalaria species are of agronomic importance, their genetic improvement is still incipient (Bhandari et al., 2016). In plant breeding programs, it is important to know the linear relationships of traits, mainly when the simultaneous selection of traits is desired, or when the main trait has low heritability, or is difficult to measure (Cruz et al., 2012). The linear relationships between traits can be evaluated with the Pearson's linear correlation coefficients (r), in the range of $-1 \leq r \leq 1$, in which the intensity of the linear correlation is larger when r is closer to |1| (Ferreira, 2009).

Complementary studies can be performed from the correlation coefficients for the definition of cause and effect relationships, and indirect selection of plants (Cruz et al., 2012). In this sense, if a given correlation matrix is estimated from an insufficient sample size, it is likely that the diagnosis of the multicollinearity by the different indicators will be biased or questionable. In addition, complementary analyses of a correlation matrix – such as partial correlation analysis, path analysis, and canonical correlation analysis – could generate biased coefficients. Also, the principal components analysis from a correlation matrix could generate biased eigenvalues and eigenvectors. Finally, any other statistical procedure, besides those mentioned, performed from an estimated correlation matrix with low precision can generate unreliable results. Therefore, if the sample size for the estimation of the correlations is insufficient, all subsequent analyses may be biased, or not compatible, with the behavior at the population level.

Given the importance of knowing the linear relations between traits, it is necessary to define the sample size to be used for the estimation of correlation coefficients. In this sense, Cargnelutti Filho et al. (2010) and Toebe et al. (2015) defined the sample size for the estimation of r in single, triple, and double corn hybrids. Toebe et al. (2015) verified that the sample size varies among corn hybrids, crops, and pairs of traits, and that a larger sample size is required to estimate the correlation coefficients between weakly correlated traits and vice-versa, in agreement with that established in studies by Bonett & Wright (2000) and Olivoto et al. (2018). The sample size to estimate the Pearson's correlation coefficients was also performed at precision levels in other agricultural crops, such as crambe (*Crambe abyssinica*) (Cargnelutti Filho et al., 2011), castor bean (*Ricinus communis*) (Cargnelutti Filho et al., 2012), and cherry tomato (*Solanum lycopersicum* 'Cerasiforme') (Sari et al., 2017). A study was recently developed to evaluate the influence of sample size and magnitude of correlation on the confidence interval width for Pearson's correlation coefficients, with real and simulated data (Olivoto et al., 2018). According to Kozak et al. (2012), it should be noted that if the correlation coefficient is estimated from a small sample size, the confidence interval for the population correlation will be very wide, and the interpretations will have little precision.

Studies of sample size for crotalaria species have been already carried out for the estimation of the mean and coefficient of variation (Toebe et al., 2017, 2018), but we did not find in the literature, studies on sample size for the estimation of correlation coefficients in this genus. It is likely that the sample size varies between species of crotalaria, and between pairs of traits of certain species.

The objective of this work was to determine the sample size necessary to estimate the Pearson's linear correlation coefficients for four species of crotalaria at precision levels.

Materials and Methods

Four uniformity trials – blank experiments, that is, without treatments – were carried out in the season of 2014/2015, in the experimental area of Universidade Federal do Pampa, campus Itaqui, located in the municipality of Itaqui (29º 09' 25" S, 56º 33' 16" W, at 74 m altitude), in the state of Rio Grande do Sul, Brazil. According to the classification of Köppen-Geiger, the climate of the region is Cfa type, humid subtropical with hot summers, and without a defined dry season (Wrege et al., 2012); its soil is classified as a Plintossolo Háplico (Santos et al., 2013), i.e., a Haplic Plinthosol.

Each one of the four species of crotalaria – *C. juncea*, *C. spectabilis*, *C. breviflora* and *C. ochroleuca* – was allocated in a uniformity trial area of 65.61 m^2 (8.1 m length × 8.1 m width), treated with fertilizer at 25 kg ha^{-1} N, 100 kg ha^{-1} P_2O_5, and 100 kg ha^{-1} K_2O.

The four species were sown in October 2014, with 0.45 m spacing between rows and 27, 33, 33, and 44 seed m of the row, respectively, for *C. juncea*, *C. spectabilis*, *C. breviflora*, and *C. ochroleuca*. The other cultural treatments were carried out in a uniform way within the sample area. In the period from March to June 2015, successive harvests of pods were held randomly, in accordance with the productive cycle of each species. From each species, 1,000 pods were collected and, in each pod, the following traits were evaluated: mass of pod with seed (MPWS), mass of pod without seed (MPWOS), length of pod (LP), width of pod (WP), height of pod (HP), number of seed per pod (NSP), mass of seed per pod (MSP = MPWS - MPWOS) and mass of one hundred seed (MHS = MSP × 100/NSP). More details of the conduction of this experiment were described by Toebe et al. (2018).

Pearson's linear correlation coefficient (r) was calculated for each species of crotalaria in the 28 pairs of traits, and the significance of r was checked out by Student's t-test, at 5% probability. The sample size was obtained via resampling with the replacement technique, which is considered adequate for conditions in which the distribution of the data is not known (Ferreira, 2009). In this sense, 199 sample sizes were planned, that is, the smallest sample size of 10 pods, and the other sample sizes obtained with the addition of five pods, in such a way that the planned sample sizes were n = 10, 15, 20, ..., 1,000 pods. For each planned sample size of each species, 10,000 resamples with replacements were obtained and, in each resample, r of each of the 28 pairs of traits were estimated. Based on the 10,000 estimates, the percentile 2.5th, the mean, and the percentile 97.5th were determined. The amplitude of the 95% confidence interval was calculated ($CI_{95\%}$) by the difference between the percentile 97.5th and the percentile 2.5th.

To determine the sample size (number of pods) required for the r estimation from each of the 28 pairs of traits, in each species, $CI_{95\%}$ of r was initially set as equal to 0.10 (higher precision), 0.20, 0.30, and 0.40 (lower precision). The optimal sample size (n) was considered as the minimum number of pods from which $CI_{95\%}$ of r was less or equal to the limit for each precision level (0.10, 0.20, 0.30 or 0.40), as previously described by Cargnelutti Filho et al. (2010), Toebe et al. (2015), and Olivoto et al. (2018). Statistical analyses were performed with the aid of the program R (R Core Team, 2018) and the Microsoft Office Excel.

Results and Discussion

Only for two (LP×MHS and WP×HP) of the 28 pairs of traits in *C. juncea*, the correlations were not

Table 1. Estimates of Pearson's linear correlation coefficients in 28 pairs of traits measured in 1,000 pods of crotalaria species – *Crotalaria juncea*, *Crotalaria spectabilis*, *Crotalaria breviflora*, and *Crotalaria ochroleuca* – in the crop year 2014/2015.

Trait pair[(1)]	*C. juncea*	*C. spectabilis*	*C. breviflora*	*C. ochroleuca*	Mean
MPWS×MPWOS	0.646*	0.892*	0.762*	0.885*	0.796
MPWS×LP	0.509*	0.690*	0.566*	0.691*	0.614
MPWS×WP	0.481*	0.736*	0.583*	0.599*	0.600
MPWS×HP	0.356*	0.472*	0.492*	0.432*	0.438
MPWS×NSP	0.773*	0.834*	0.779*	0.654*	0.760
MPWS×MSP	0.984*	0.927*	0.942*	0.915*	0.942
MPWS×MHS	0.497*	0.294*	0.442*	0.555*	0.447
MPWOS×LP	0.685*	0.701*	0.699*	0.705*	0.697
MPWOS×WP	0.571*	0.760*	0.649*	0.692*	0.668
MPWOS×HP	0.390*	0.449*	0.578*	0.483*	0.475
MPWOS×NSP	0.390*	0.586*	0.372*	0.395*	0.436
MPWOS×MSP	0.501*	0.657*	0.499*	0.621*	0.569
MPWOS×MHS	0.265*	0.220*	0.300*	0.436*	0.305
LP×WP	0.454*	0.686*	0.494*	0.590*	0.556
LP×HP	0.395*	0.304*	0.412*	0.519*	0.407
LP×NSP	0.435*	0.558*	0.332*	0.412*	0.434
LP×MSP	0.418*	0.569*	0.394*	0.551*	0.483
LP×MHS	0.043^{ns}	0.067*	0.167*	0.312*	0.147
WP×HP	0.052^{ns}	0.386*	0.522*	0.489*	0.362
WP×NSP	0.254*	0.572*	0.331*	0.267*	0.356
WP×MSP	0.413*	0.597*	0.442*	0.406*	0.464
WP×MHS	0.314*	0.110*	0.265*	0.274*	0.241
HP×NSP	0.319*	0.366*	0.292*	0.182*	0.290
HP×MSP	0.313*	0.415*	0.358*	0.308*	0.348
HP×MHS	0.073*	0.147*	0.170*	0.240*	0.157
NSP×MSP	0.787*	0.904*	0.848*	0.759*	0.825
NSP×MHS	-0.105*	-0.118*	-0.093*	-0.106*	-0.105
MSP×MHS	0.502*	0.308*	0.435*	0.556*	0.450
Correlation between species calculated with the correlation of the 28 pairs of traits					
	CJ	CS	CB	CO	
C. juncea (CJ)	-	0.842	0.893	0.858	
C. spectabilis (CS)	0.842	-	0.884	0.820	
C. breviflora (CB)	0.893	0.884	-	0.922	
C. ochroleuca (CO)	0.858	0.820	0.922	-	

[(1)]MPWS, mass of pods with seed; MPWOS, mass of pods without seed; LP, length of pod; WP, width of pod; HP, height of pod; NSP, number of seed per pod; MSP, mass of seed per pod; MHS, mass of 100 seed. *Significant by the t-test, at 5% probability. [ns]Nonsignificant.

significant (Table 1). In *C. spectabilis*, *C. breviflora*, and *C. ochroleuca* all trait pairs showed significant correlations. Thus, out of the 112 evaluated correlation (4 species × 28 pairs of traits), 110 were significant at 5% probability. It is important to observe the practical significance, since the high original sample size (1,000 pods) causes low-magnitude correlations to become significant. As highlighted by Hair Jr. et al. (2009), the practical significance indicates whether the result is useful or not to achieve the research objectives. In this sense, Mukaka (2012) emphasizes that the misuse of correlation is common among researchers. According to Kozak (2008) and Kozak et al. (2012), very small correlation coefficients can be statistically significant, when a large sample size is used and vice-versa. According to these authors, significance merely suggests the presence of a nonzero population correlation coefficient, not necessarily an important correlation.

Adopting the classification of the correlation coefficient with practical magnitude proposed by Hinkle et al. (2003), in all species of crotalaria, the correlation between MPWS×MSP was very high (0.90 to 1.00) (Table 1). The very high correlation between these two variables is expected, since MSP is obtained from the difference between MPWS and MPWOS, that is, the smaller the MPWOS interference, the greater the association between MSP and MPWS. In *C. spectabilis*, a very high correlation was also observed between NSP×MSP. A high and positive correlation (0.70 to 0.90) was found between the following trait pairs: MPWS×NSP and NSP×MSP in *C. juncea*; MPWS×MPWOS, MPWS×WP, MPWS×NSP, MPWOS×LP, and MPWOS×WP, in *C. spectabilis*; MPWS×MPWOS, MPWS×NSP, and NSP×MSP, in *C. breviflora*; and between MPWS×MPWOS, MPWOS×LP, and NSP×MSP, in *C. ochroleuca* (Figures 1 and 2).

Correlations considered negligible from a practical point of view ($-0.30 \leq r \leq 0.30$) were obtained for the following trait pairs: MPWOS×MHS, LP×MHS, WP×HP, WP×NSP, HP×MHS, and NSP×MHS, in *C. juncea*; for MPWS×MHS, MPWOS×MHS, LP×MHS, WP×MHS, HP×MHS, and NSP×MHS, in *C. spectabilis*; for MPWOS×MHS, LP×MHS, WP×MHS, HP×NSP, HP×MHS, and NSP×MHS, in *C. breviflora*; and for WP×NSP, WP×MHS, HP×NSP, HP×MHS, and NSP×MHS, in *C. ochroleuca* (Table 1). In general, MHS showed the lowest values of correlation with the other traits. The other pairs of traits showed low or moderate positive correlations (0.30 to 0.70). Additionally, Pearson's linear correlation coefficients between species, based on the 28 correlation values between pairs of traits, were high to very high ($0.842 \leq r \leq 0.922$), indicating that, in general, the studied crotalaria species have similar association patterns.

Depending on the pair of traits considered, the sample size for the estimation of the Pearson's linear correlation coefficient with the highest precision, established in this study ($CI_{95\%}$ of 0.10), ranged as follows: from 10 to more than 1,000 pods in *C. juncea*; from 45 to more than 1,000 pods in *C. spectabilis*; from 25 to more than 1,000 pods in *C. breviflora*; and from 50 to more than 1,000 pods in *C. ochroleuca* (Table 2). For all species, the smallest sample size at this level of precision was verified for the correlation between MPWS×MSP. As previously mentioned, this pair of traits was the only one to show a very high correlation (Table 1), according to the classification of Hinkle et al. (2003), in all species of crotalaria, as expected, since MSP is obtained from the difference between MPWS and MPWOS. These results indicate that high correlations can be estimated with precision from smaller sample sizes.

Considering an intermediate precision in the Pearson's linear correlation coefficient estimation ($CI_{95\%}$ of 0.20), the sample size ranged from 10 to 440 pods in *C. juncea*, from 15 to 415 pods in *C. spectabilis*, from 10 to 425 pods in *C. breviflora*, and from 20 to 380 pods in *C. ochroleuca*, depending on the pair of traits considered (Table 2). In general, a larger magnitude of correlations was found for MPWS×MPWOS, MPWS×NSP, MPWS×MSP, and NSP×MSP; in at least three species, these correlations were considered high or very high (Table 1). Accordingly, in general, these pairs of traits required the smallest sample size for the estimation of correlations (Table 2). However, in at least three species, the correlations between MPWS×MHS, LP×MHS, WP×MHS, HP×MHS, and NSP×MHS were considered negligible. In these pairs of traits, in general, a larger sample size was required to estimate the correlations. The use of 440 pods would allow of the estimation of correlations with 0.20 as the maximum $CI_{95\%}$, independently of the species and pair of traits considered. Thus, if, for instance, an experiment with five treatments and four replicates is carried out with 20 plots, the evaluations should be performed for 22

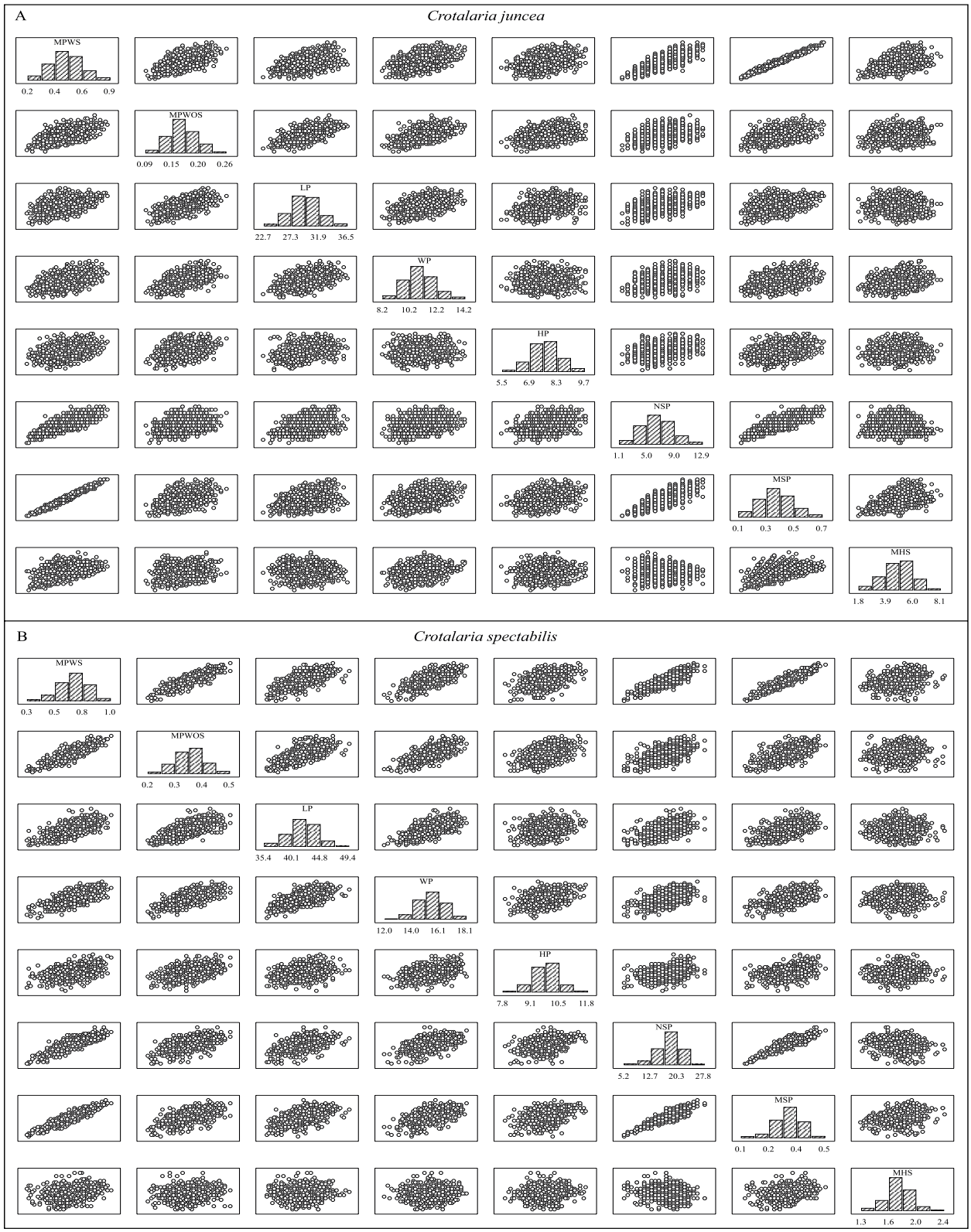

Figure 1. Matrix with a histogram frequency (in diagonal) and dispersion graphs between mass of pod with seed (MPWS, g), mass of pods without seed (MPWOS, g), length of pod (LP, mm), width of pod (WP, mm), height of pod (HP, mm), number of seed per pod (NSP, unity), mass of seed per pod (MSP, g), and mass of 100 seed (MHS, g), in 1,000 pods of *Crotalaria juncea* (A) and *Crotalaria spectabilis* (B).

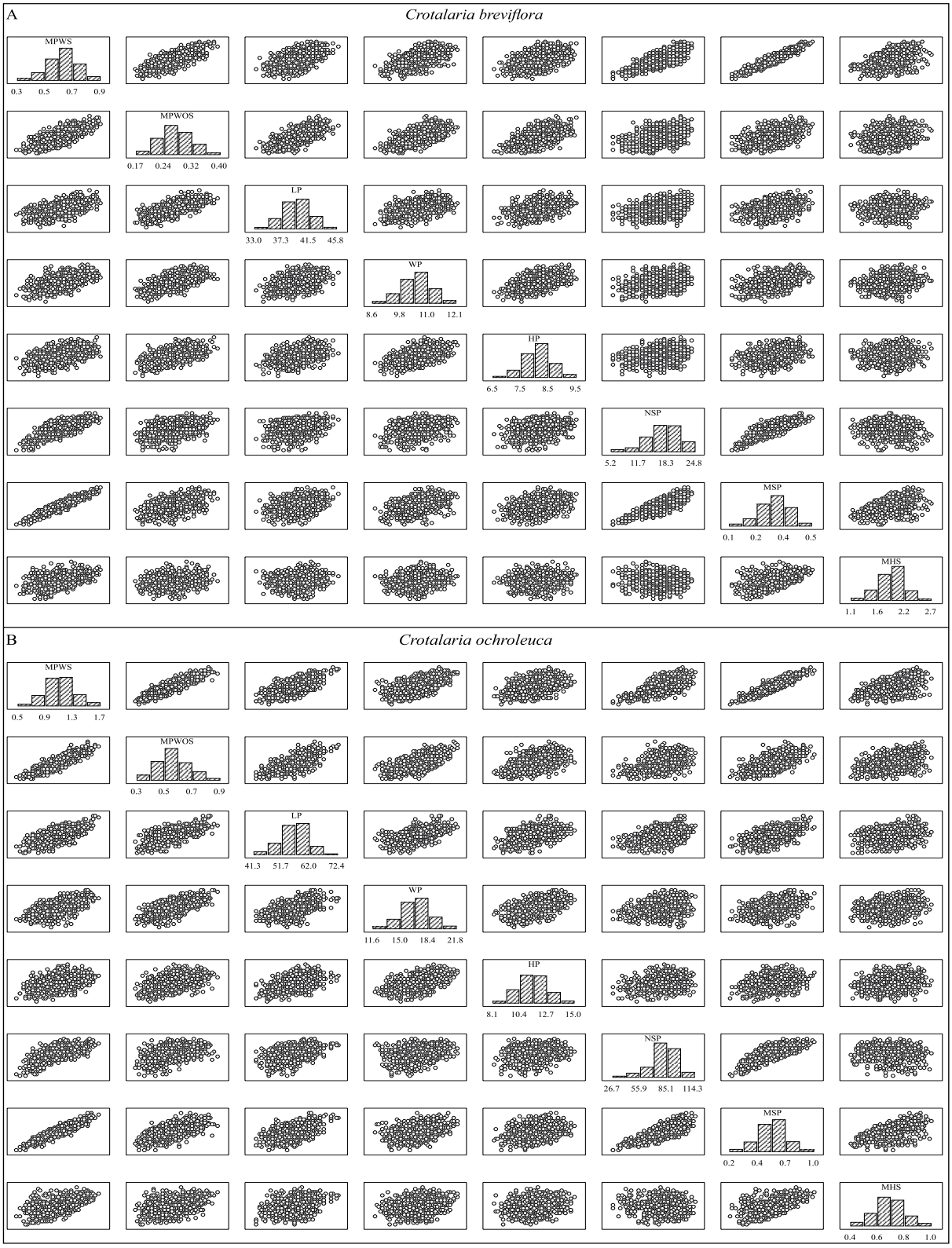

Figure 2. Matrix with a histogram frequency (in diagonal) and dispersion graphs between mass of pod with seed (MPWS, g), mass of pods without seed (MPWOS, g), length of pod (LP, mm), width of pod (WP, mm), height of pod (HP, mm), number of seed per pod (NSP, unity), mass of seed per pod (MSP, g), and mass of 100 seed (MHS, g), in 1,000 pods of Crotalaria breviflora (a) and Crotalaria ochroleuca (b).

pods per plot to estimate the correlation at this precision level. That is, the evaluation of 22 pods per plot would allow to adequately estimate the correlation of all pairs of traits, irrespectively of the crotalaria species used, with an executable number of measurements from a practical point of view.

As previously mentioned, in all species, the correlation between MPWS×MSP was very high, and the correlation between NSP×MHS was negligible from a practical point of view (Table 1). In this sense, it is possible to verify the difference of the confidence interval of the correlation coefficients for these two pairs of traits in all the species (Figure 3 A-H). Also, it can be verified that the sample size required to estimate the linear correlations decreases as the correlation strength increases (Figure 4). In this sense, Olivoto et al. (2018) verified that the Pearson's confidence interval width is inversely proportional to the strength of the

Table 2. Sample size (number of pods) for the estimation of Pearson's linear correlation coefficients in 28 pairs of traits, in *Crotalaria juncea*, *Crotalaria spectabilis*, *Crotalaria breviflora*, and *Crotalaria ochroleuca*), with 95% confidence interval ($CI_{95\%}$) of 0.10 and 0.20.

Trait pair[(1)]	$CI_{95\%}$ of 0.10				$CI_{95\%}$ of 0.20			
	C. juncea	*C. spectabilis*	*C. breviflora*	*C. ochroleuca*	*C. juncea*	*C. spectabilis*	*C. breviflora*	*C. ochroleuca*
MPWS×MPWOS	545	90	295	75	145	30	80	25
MPWS×LP	915	470	630	470	235	120	165	120
MPWS×WP	865	335	670	675	220	90	170	175
MPWS×HP	>1,000	>1,000	810	955	270	275	205	240
MPWS×NSP	265	175	250	550	70	50	70	140
MPWS×MSP	10	45	25	50	10	15	10	20
MPWS×MHS	845	>1,000	>1,000	>1,000	215	355	305	260
MPWOS×LP	440	465	330	395	115	120	90	110
MPWOS×WP	605	310	460	430	150	85	120	110
MPWOS×HP	>1,000	995	615	915	270	245	160	225
MPWOS×NSP	>1,000	780	>1,000	>1,000	290	200	280	275
MPWOS×MSP	915	675	935	575	240	175	240	150
MPWOS×MHS	>1,000	>1,000	>1,000	>1,000	375	395	320	310
LP×WP	885	480	945	600	230	125	240	155
LP×HP	>1,000	>1,000	930	815	335	340	240	210
LP×NSP	>1,000	800	>1,000	>1,000	280	210	315	255
LP×MSP	>1,000	770	>1,000	800	305	200	265	210
LP×MHS	>1,000	>1,000	>1,000	>1,000	370	355	330	345
WP×HP	>1,000	>1,000	755	875	440	320	195	220
WP×NSP	>1,000	735	>1,000	>1,000	350	195	335	365
WP×MSP	>1,000	675	>1,000	>1,000	260	175	265	300
WP×MHS	>1,000	>1,000	>1,000	>1,000	265	340	270	330
HP×NSP	>1,000	>1,000	>1,000	>1,000	315	325	330	370
HP×MSP	>1,000	>1,000	>1,000	>1,000	285	310	280	315
HP×MHS	>1,000	>1,000	>1,000	>1,000	315	350	365	360
NSP×MSP	250	85	155	305	70	30	45	80
NSP×MHS	>1,000	>1,000	>1,000	>1,000	370	415	425	380
MSP×MHS	790	>1,000	>1,000	>1,000	205	390	335	265

[(1)]MPWS, mass of pods with seed; MPWOS, mass of pods without seed; LP, length of pod; WP, width of pod; HP, height of pod; NSP, number of seed per pod; MSP, mass of seed per pod; MHS, mass of 100 seed.

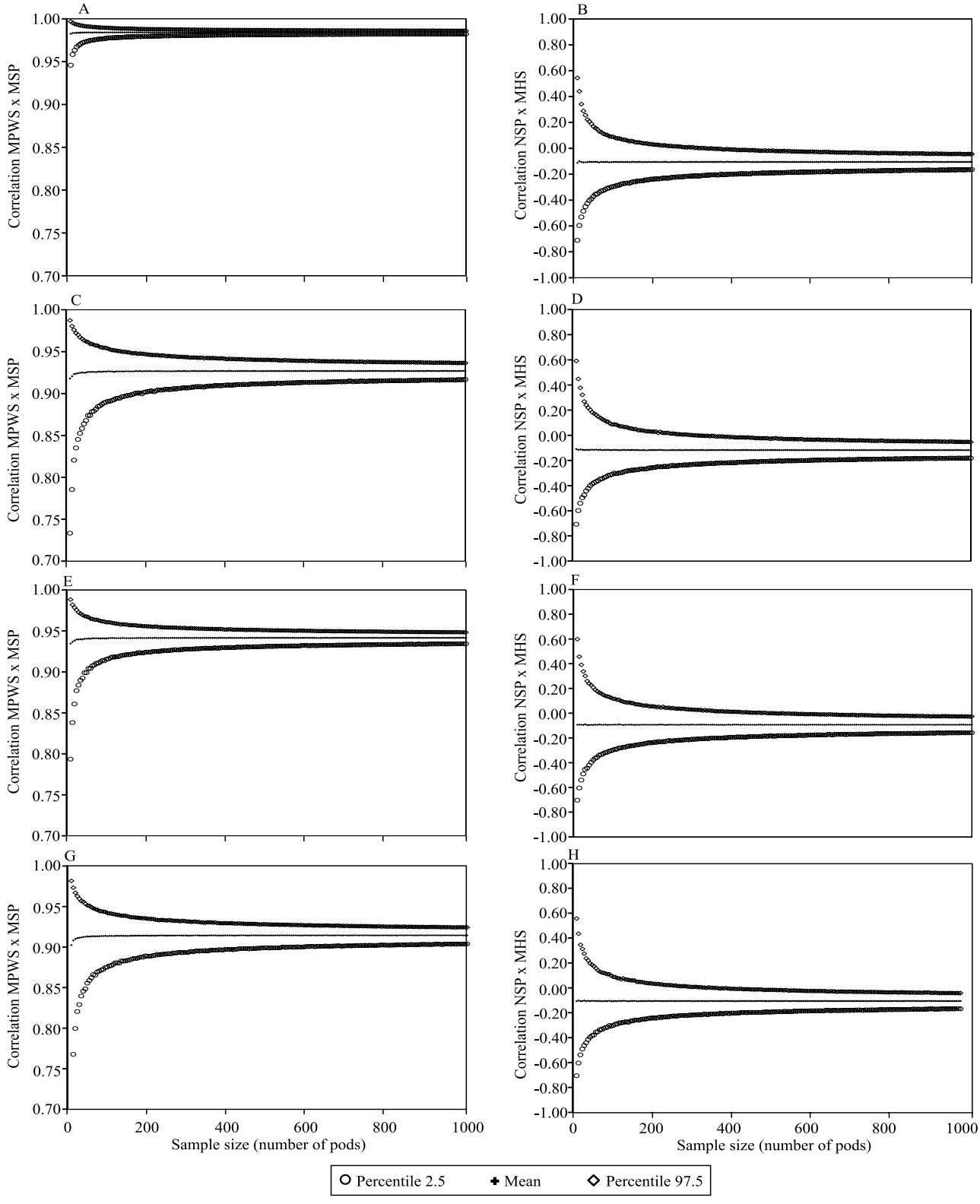

Figure 3. Percentile 2.5th, mean, and percentile 97.5th of 10,000 estimates of Pearson's linear correlation coefficients, as follows: between mass of pod with seed and mass of seed per pod (MPWS×MSP) in *Crotalaria juncea* (A), *Crotalaria spectabilis* (C); *Crotalaria breviflora* (E); *Crotalaria ochroleuca* (G); and between number of seed per pod and mass of 100 seed (NSP×MHS) in *C. juncea* (B), *C. spectabilis* (D), *C. breviflora* (F), and *C. ochroleuca* (H).

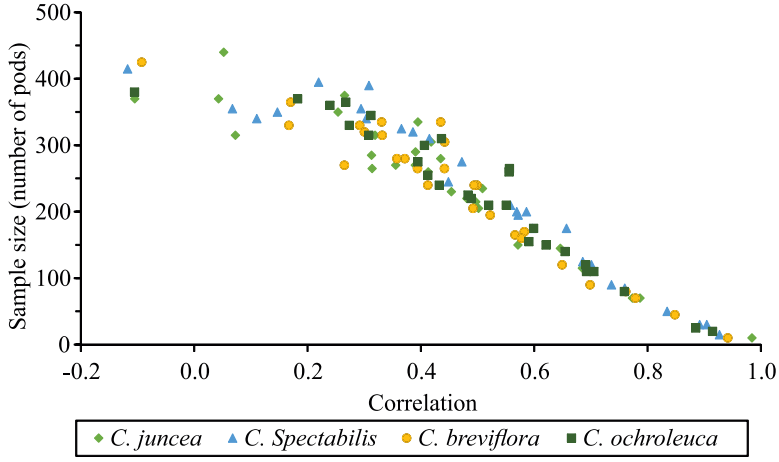

Figure 4. Relationships between the correlation magnitude and the sample size recommended for *Crotalaria juncea*, *Crotalaria spectabilis*, *Crotalaria breviflora*, and *Crotalaria ochroleuca*), based on the 28 correlations of the trait pairs of each species (Table 1) and the corresponding sample size for $CI_{95\%}$ of 0.20 (Table 2).

Table 3. Sample size (number of pods) for the estimation of Pearson's linear correlation coefficients in 28 pairs of traits, in *Crotalaria juncea*, *Crotalaria spectabilis*, *Crotalaria breviflora*, and *Crotalaria ochroleuca*, with 95% confidence interval ($CI_{95\%}$) of 0.30 and 0.40.

Trait pair[(1)]	$CI_{95\%}$ of 0.30				$CI_{95\%}$ of 0.40			
	C. juncea	*C. spectabilis*	*C. breviflora*	*C. ochroleuca*	*C. juncea*	*C. spectabilis*	*C. breviflora*	*C. ochroleuca*
MPWS×MPWOS	70	15	40	15	40	10	25	10
MPWS×LP	110	60	75	60	60	35	45	35
MPWS×WP	100	45	75	80	55	25	45	50
MPWS×HP	125	120	95	110	70	70	55	65
MPWS×NSP	35	25	35	65	20	15	25	40
MPWS×MSP	10	10	10	10	10	10	10	10
MPWS×MHS	95	160	135	120	55	90	75	70
MPWOS×LP	55	55	40	50	35	35	25	30
MPWOS×WP	70	40	60	50	40	25	35	30
MPWOS×HP	120	115	75	105	70	65	45	60
MPWOS×NSP	135	90	130	130	75	55	75	70
MPWOS×MSP	105	85	110	65	60	45	65	40
MPWOS×MHS	175	185	150	135	95	105	85	80
LP×WP	105	60	110	70	60	35	65	40
LP×HP	155	155	110	95	85	90	65	55
LP×NSP	125	95	140	120	75	55	85	70
LP×MSP	135	85	115	95	80	55	70	55
LP×MHS	160	160	150	155	95	95	85	95
WP×HP	200	140	90	100	110	80	50	60
WP×NSP	155	90	150	165	85	55	90	95
WP×MSP	120	80	120	135	70	45	70	80
WP×MHS	115	150	125	150	70	90	70	85
HP×NSP	140	145	150	165	80	85	85	100
HP×MSP	130	140	135	140	70	80	75	80
HP×MHS	150	160	160	160	85	95	95	95
NSP×MSP	35	15	25	40	20	10	15	25
NSP×MHS	160	185	190	180	95	100	110	105
MSP×MHS	95	175	155	120	55	95	90	70

[(1)]MPWS, mass of pods with seed; MPWOS, mass of pods without seed; LP, length of pod; WP, width of pod; HP, height of pod; NSP, number of seed per pod; MSP, mass of seed per pod; and MHS, mass of 100 seed.

association between traits. The inverse relationship of strength of association between traits and sample sizes needed to estimate the correlations was also observed in studies applied to maize (Cargnelutti Filho et al., 2010; Toebe et al., 2015), crambe (Cargnelutti Filho et al., 2011), castor bean (Cargnelutti Filho et al., 2012) and cherry tomato (Sari et al., 2017).

Considering $CI_{95\%}$ of 0.30, the sample size ranged from 10 to 200 pods in *C. juncea*, from 10 to 185 pods in *C. spectabilis*, from 10 to 190 pods in *C. breviflora*, and from 10 to 180 pods in *C. ochroleuca*, depending on the pair of traits considered (Table 3). Considering $CI_{95\%}$ of 0.40, the sample size ranged from 10 to 110 pods in *C. juncea*, from 10 to 105 pods in *C. spectabilis*, from 10 to 110 pods in *C. breviflora*, and from 10 to 105 pods in *C. ochroleuca*, depending on the pair of traits considered. In general, a greater variability of sample size was observed between the pairs of traits than between species for a given pair of traits (Tables 2 and 3).

Conclusions

1. The sample size varies between crotalaria species and, especially, between pairs of traits as a function of the magnitude of the correlation coefficient.

2. Smaller sample sizes are required to estimate the correlation coefficients between highly correlated traits.

3. To estimate the correlation coefficients with $CI_{95\%}$ of 0.20, 10 to 440 pods are required, depending on the species, pairs of traits, and magnitude of the correlation coefficient.

Acknowledgments

To Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (Fapergs), the Programa de Educação Tutorial do Ministério da Educação, and to Fundação Universidade Federal do Pampa (Unipampa), for scholarships and financial support; to the scholarship students and volunteers, for their help in the conduction of the experiment and data collection, and to the company Piraí Sementes, for granting seed for research purposes.

References

BHANDARI, H.R.; TRIPATHI, M.K.; CHAUDHARY, B.; SARKAR, S.K. Sunnhemp breeding: Challenges and prospects. **Indian Journal of Agricultural Sciences**, v.86, p.1391-1398, 2016.

BONETT, D.G.; WRIGHT, T.A. Sample size requirements for Pearson, Kendall, and Spearman correlations. **Psychometrika**, v.65, p.23-28, 2000. DOI: https://doi.org/10.1007/BF02294183.

BRAZ, G.B.P.; OLIVEIRA JR., R.S. de; CONSTANTIN, J.; RAIMONDI, R.T.; RIBEIRO, L.M.; GEMELLI, A.; TAKANO, H.K. Plantas daninhas como hospedeiras alternativas para *Pratylenchus brachyurus*. **Summa Phytopathologica**, v.42, p.233-238, 2016. DOI: https://doi.org/10.1590/0100-5405/2129.

CARGNELUTTI FILHO, A.; LOPES, S.J.; BRUM, B.; TOEBE, M.; SILVEIRA, T.R. da; CASAROTTO, G. Tamanho de amostra para a estimação do coeficiente de correlação linear de Pearson entre caracteres de mamoneira. **Semina: Ciências Agrárias**, v.33, p.953-962, 2012. DOI: https://doi.org/10.5433/1679-0359.2012v33n3p953.

CARGNELUTTI FILHO, A.; LOPES, S.J.; TOEBE, M.; SILVEIRA, T.R. da; SCHWANTES, I.A. Tamanho de amostra para estimação do coeficiente de correlação de Pearson entre caracteres de *Crambe abyssinic*a. **Revista Ciência Agronômica**, v.42, p.149-158, 2011. DOI: https://doi.org/10.1590/S1806-66902011000100019.

CARGNELUTTI FILHO, A.; TOEBE, M.; BURIN, C.; SILVEIRA, T.R. da; CASAROTTO, G. Tamanho de amostra para estimação do coeficiente de correlação linear de Pearson entre caracteres de milho. **Pesquisa Agropecuária Brasileira**, v.45, p.1363-1371, 2010. DOI: https://doi.org/10.1590/S0100-204X2010001200005.

CHAUDHARY, B. Traditional cultivation of sunnhemp (*Crotalaria juncea*) in eastern India. **Indian Journal of Agricultural Sciences**, v.86, p.369-372, 2016.

CRUZ, C.D.; REGAZZI, A.J.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. 4.ed. Viçosa: UFV, 2012. v.1, 514p.

DEBERDT, P.; GOZÉ, E.; CORANSON-BEAUDU, R.; PERRIN, B.; FERNANDES, P.; LUCAS, P.; RATNADASS, A. *Crotalaria spectabilis* and *Raphanus sativus* as previous crops show promise for the control of bacterial wilt of tomato without reducing bacterial populations. **Journal of Phytopathology**, v.163, p.377-385, 2015. DOI: https//.doi.org/10.1111/jph.12333.

DINIZ, E.R.; VARGAS, T.O.; PEREIRA, W.D.; SANTOS, R.H.S.; URQUIAGA, S.; MODOLO, A.J. Levels of *Crotalaria juncea* on growth, production, recovery and efficiency of the use of N in broccoli. **Horticultura Brasileira**, v.35, p.395-401, 2017. DOI: https://doi.org/10.1590/s0102-053620170313.

ELSAID, E.; SILVA, R. Potential of sun hemp residue to provide potato with adequate nitrogen. **Journal of Plant Nutrition**, v.40, p.851-860, 2017. DOI: https://doi.org/10.1080/01904167.2016.1262397.

FERREIRA, D.F. **Estatística básica**. 2.ed. rev. Lavras: UFLA, 2009. 664p.

HAIR JR., J.F.; BLACK, W.C.; BABIN, B.J.; ANDERSON, R.E.; TATHAM, R.L. **Análise multivariada de dados**. 6.ed. São Paulo: Bookman, 2009. 688p.

HINKLE, D.E.; WIERSMA, W.; JURS, S.G. **Applied Statistics for the Behavioral Sciences**. 5th ed. Boston: Houghton Mifflin , 2003. 756p.

KOZAK, M. Correlation coefficient and the fallacy of statistical hypothesis testing. **Current Science**, v.95, p.1121-1122, 2008.

KOZAK, M.; KRZANOWSKI, W.; TARTANUS, M. Use of the correlation coefficient in agricultural sciences: problems, pitfalls and how to deal with them. **Anais da Academia Brasileira de Ciências**, v.84, p.1147-1156, 2012. DOI: https://doi.org/10.1590/S0001-37652012000400029.

MUKAKA, M.M. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. **Malawi Medical Journal**, v.24, p.69-71, 2012.

OLIVOTO, T.; LÚCIO, A.D.C.; SOUZA, V.Q. de; NARDINO, M.; DIEL, M.I.; SARI, B.G.; KRYSCZUN, D.K.; MEIRA, D.; MEIER, C. Confidence interval width for Pearson's correlation coefficient: a Gaussian-independent estimator based on sample size and strength of association. **Agronomy Journal**, v.110, p.503-510, 2018. DOI: https://doi.org/10.2134/agronj2017.09.0566.

R Core Team. **R**: a language and environment for statistical computing. 2018. Available at: <https://www.r-project.org/>. Accessed on: May 5 2018.

REIGADA, C.; GUIMARÃES, K.F.; PARRA, J.R.P. Relative fitness of *Helicoverpa armigera* (Lepidoptera: Noctuidae) on seven host plants: a perspective for IPM in Brazil. **Journal of Insect Science**, v.16, p.1-5, 2016. DOI: https://doi.org/10.1093/jisesa/iev158.

SANTOS, H.G. dos; JACOMINE, P.K.T.; ANJOS, L.H.C. dos; OLIVEIRA, V.A. de; LUMBRERAS, J.F.; COELHO, M.R.; ALMEIDA, J.A. de; CUNHA, T.J.F.; OLIVEIRA, J.B. de. **Sistema brasileiro de classificação de solos**. Brasília: Embrapa, 2013. 353p.

SARI, B.G.; LÚCIO, A.D.; SANTANA, C.S.; KRYSCZUN, D.K.; TISCHLER, A.L.; DREBES, L. Sample size for estimation of the Pearson correlation coefficient in cherry tomato tests. **Ciência Rural**, v.47, p.1-6, 2017. DOI: https://doi.org/10.1590/0103-8478cr20170116.

TOEBE, M.; BANDEIRA, C.T.; FORTES, S.K.G.; CARVALHO, J.O. de; TARTAGLIA, F. de L.; TAMBARA, A.L.; MELO, P.J. de. Dimensionamento amostral e associação linear entre caracteres de *Crotalaria spectabilis*. **Bragantia**, v.76, p.45-53, 2017. DOI: https://doi.org/10.1590/1678-4499.653.

TOEBE, M.; CARGNELUTTI FILHO, A.; LOPES, S.J.; BURIN, C.; SILVEIRA, T.R. da; CASAROTTO, G. Sample size in the estimation of correlation coefficients for corn hybrids in crops and accuracy levels. **Bragantia**, v.74, p.16-24, 2015. DOI: https://doi.org/10.1590/1678-4499.0324.

TOEBE, M.; MACHADO, L.N.; TARTAGLIA, F.L.; CARVALHO, J.O. de; BANDEIRA, C.T.; CARGNELUTTI FILHO, A. Sample size for estimating mean and coefficient of variation in species of crotalarias. **Anais da Academia Brasileira de Ciências**, v.90, p.1705-1715, 2018. DOI: https://doi.org/10.1590/0001-3765201820170813.

WREGE, M.S.; STEINMETZ, S.; REISSER JÚNIOR, C.; ALMEIDA, I.R. de. (Ed.). **Atlas climático da Região Sul do Brasil**: Estados do Paraná, Santa Catarina e Rio Grande do Sul. 2.ed. Brasília: Embrapa, 2012. 333p.