

<http://dx.doi.org/10.15361/1984-5529.2019v47n3p321-326>

Discrimination of forage pea seed lots by means of multivariate techniques

Discriminação de lotes de sementes de ervilha-forrageira por meio de técnicas multivariadas

Carla Gomes MACHADO¹; Cibele Chalita MARTINS²; Givanildo Zildo da SILVA³; Simério Carlos Silva CRUZ¹
Gabriela Fernandes GAMA⁴; Mirelle Vaz COELHO⁴

¹ Engenheiros Agrônomos, Professores, Doutores, Universidade Federal de Goiás - UFG, Campus Jatobá, Jataí, GO, Brasil. carlagomesmachado@gmail.com; simerio_cruz@yahoo.com.br

² Engenheira Agrônoma, Professora Livre-Docente, Faculdade de Ciências Agrárias e Veterinárias - UNESP, Jaboticabal, SP, Brasil. cibele.chalita@unesp.br

³ Autor para correspondência. Engenheiro Agrônomo, Professor, Universidade Federal de Goiás - UFG, Campus Jatobá, Jataí, GO, Brasil. givanildozildo@hotmail.com

⁴ Engenheira Agrônoma e Bióloga, Mestrandas, Universidade Federal de Goiás - UFG, Campus Jatobá, Jataí, GO, Brasil. gabifgama@hotmail.com; mirellevaz7@gmail.com

Recebido em: 25-03-2019; Aceito em: 20-05-2019

Abstract

Multivariate techniques allow to understand the structural dependence contained in the variables, as well as to characterize groups of seed lots according to specific standards. Thus, this study analyzes the efficiency of multivariate exploratory techniques in discriminating forage pea seed lots as a function of the physiological potential of seeds. We evaluated ten seed lots of forage pea in a completely randomized design, considering the following variables: thousand seed weight, germination, first germination count, electrical conductivity, and accelerated aging. Moreover, seedling emergence, first count of seedlings in the field, and seedling emergence speed index in the field were added to randomized blocks with four replications per lot. Initially, the data obtained in each test were analyzed separately by means of analysis of variance, and the means of the treatments were compared by the Scott Knott test at 5% probability. Exploratory multivariate statistical techniques were applied by means of Cluster Analysis and Principal Components Analysis to discriminate seed lots with better physiological quality and to characterize the variables responsible for the differentiation between them. Multivariate analysis of principal components is efficient in discriminating vigor and seed germination tests in *Pisum sativum* subsp. *Arvense*, which help in identifying lots of superior performance in the field.

Additional keywords: cluster analysis; germination; *Pisum sativum* subsp. *Arvense*; principal component analysis; vigor.

Resumo

Técnicas multivariadas permitem compreender a dependência estrutural contida nas variáveis, bem como caracterizar grupos de lotes de sementes segundo padrões específicos. Assim, objetivou-se verificar a eficiência de técnicas exploratórias multivariadas para discriminar lotes de sementes de ervilha-forrageira em função do potencial fisiológico. Foram utilizados dez lotes de sementes de ervilha-forrageira, avaliados em delineamento inteiramente casualizado para massa de mil sementes, germinação, primeira contagem de germinação, condutividade elétrica e envelhecimento acelerado. A emergência, a primeira contagem e o índice de velocidade de plântulas em campo foram implantados em blocos casualizados, com quatro repetições por lote. Inicialmente, os dados obtidos em cada teste foram analisados separadamente por meio de análise de variância, e as médias dos tratamentos foram comparadas pelo teste de Scott Knott, a 5% de probabilidade. Para discriminar os lotes de sementes com melhor qualidade fisiológica e para caracterizar as variáveis responsáveis pela diferenciação entre lotes de sementes, as técnicas exploratórias de estatística multivariada foram aplicadas por meio da Análise de Cluster e Análise de Componentes Principais. A análise multivariada de componentes principais é eficiente para discriminar os testes de vigor e a germinação de sementes *Pisum sativum* subsp. *Arvense* capazes de identificar os lotes de desempenho superior em campo.

Palavras-chave adicionais: análise de cluster; análise de componentes principais; germinação; *Pisum sativum* subsp. *Arvense*; vigor.

Introduction

Forage pea (*Pisum sativum* subsp. *arvense* L.) is a legume grown in Brazil during winter. This plant is indicated for green fertilization, soil cover, animal feed, and is presented as an advantageous alternative when compared to other cover crops due to rapid initial growth, precocity, and uniformity (Cargnelutti Filho et al., 2015).

The use of high quality seeds is essential for the rapid establishment of seedlings in the field, providing a uniform stand (Pinto et al., 2016). The germination test is routinely used by the seed industry to determine physiological quality, and the results of vigor analyses are used to identify differences in the physiological quality of the lots, especially those with a similar percentage of normal seedlings (Brasil, 2009). Such analyses, performed in seed analysis laboratories, result in several partial information that aim to evaluate the physiological quality of the lots (Lorentz & Nunes, 2013). This information can help in the internal decision making of seed producing companies as to the destination of a given lot (Marcos Filho, 2015).

In this series of tests, results are likely to be redundant, little informative, or ineffective in discriminating seed lot quality. This is because they are highly related to other characteristics. Low discriminatory capacity highlights the identification and elimination of variables that do not present additional information to the user (Lorentz & Nunes, 2013), which can be done through multivariate analysis. This analysis represents a set of exploratory techniques that allows to discriminate seed lots with common characteristics, and aims at the minimum numbers needed to describe relevant information contained in the original observations (Barbosa et al., 2013).

Groups or clusters are created for internal homogeneity to prevail. In this sense, this set of techniques, also known as cluster analysis, aims mainly at allocating observations in a relatively small number of clusters internally homogeneous and heterogeneous with each other and that represent the joint behavior of observations from certain variables (Fávero & Belfiore, 2015).

Exploratory multivariate statistical techniques are usually applied in studies with genotype assessment, such as in cowpea (Santos et al., 2016) or in fruit-bearing species, as is the case of bacurizeiro (Souza et al., 2016). However, this technique has potential for seed quality assessment, a fact proven for soybean (Barbosa et al., 2013) and *Brachiaria brizantha* cv. BRS Piatã (Silva et al., 2017). These authors state that factors such as number of variables, the need to improve the understanding of the factors influencing the physiological potential of seeds, and the accuracy of the discrimination between seed lots makes the use of multivariate techniques possible and effective in seed studies.

Thus, this study analyzes the efficiency of multivariate exploratory techniques in discriminating forage pea seed lots as a function of the physiological potential of seeds.

Materials and methods

The present study was carried out in the Laboratory of Seed Analysis of the Department of Plant Production of the Faculty of Agronomic Sciences of the São Paulo State University "Júlio de Mesquita Filho" - FCA/UNESP, Botucatu Campus, SP. Ten lots of forage pea seeds were used, provided by the Agronomic Institute of Paraná - IAPAR, Londrina-PR. The newly collected samples were homogenized and then packed in Kraft paper bags in an air-conditioned environment at 5 °C.

For the purposes of comparison and differentiation of the physiological quality of the lots, the seeds were subjected to the following determinations and evaluations:

Moisture content - determined before and after the accelerated aging tests, using two subsamples of 20 seeds for each lot, through the oven method at 105 ± 3 °C for 24 hours (Brasil, 2009).

Thousand seed weight - carried out with eight subsamples of 100 seeds per lot (Brasil, 2009).

Seedling emergence in the field - conducted in flower beds inside a screened greenhouse, with plots consisting of two furrows of 1.5 m, spaced 0.2 m apart, with 50 seeds per furrow being sown at depths of 4 to 5 cm. The counts were performed daily up to 14 days after sowing, when seedling emergence percentage in the field was calculated according to Machado et al. (2011).

First count of seedling emergence in the field: seedlings were counted when they represented $50 + 1\%$ of the final emergence. Seedling emergence speed in the field: calculated from the data of daily counts of seedlings emerged in the emergence test, and determined by the formula proposed by Maguire (1962).

Germination - carried out on paper roll substrate, moistened with an amount of water equivalent to 2.5 times the weight of the dry substrate and incubated at 20 °C, determining the percentage of normal and abnormal seedlings on the seventh day after sowing (Machado et al., 2011). First germination count: performed together with the germination test, on the fourth day after sowing (Machado et al., 2011).

Electrical conductivity - 50 seeds were placed in plastic cups to soak with 250 mL of distilled water at 25 °C for 24 hours (Machado et al., 2011). The electrical conductivity of the solution was determined by a conductivity meter and the results were expressed in $\mu\text{S cm}^{-1} \text{g}^{-1}$.

Accelerated aging - 220 seeds were used for each lot, distributed in a single uniform layer on an aluminum screen, which was inserted in plastic boxes (11.0 x 11.0 x 3.5 cm) suitable for the accelerated aging test. A saturated solution with 87% relative humidity (40 mL of saturated solution of potassium chloride 32 g KCl /100 mL of distilled water) was added at the bottom of each plastic box, and the enclosed boxes were packed in plastic bags and kept at 42 °C for a

48-hour exposure period. Seed vigor was determined after each treatment through the germination test, with counts performed on the fourth day and determination of seed moisture content (Machado et al., 2019; Brasil, 2009).

Exploratory multivariate statistical techniques were applied through Cluster Analysis and Principal Components Analysis to discriminate seed lots with better physiological quality and to characterize the variables responsible for the differentiation between them.

The experimental design was completely randomized, with the exception of seedling emergence in the field, using randomized blocks with four replications per treatment. Multivariate analyses were only performed after standardization of the variables, each with mean 0 and variance 1. Subsequently, cluster analysis was processed by calculating the Euclidean distance between the lots for the set of eight variables and using Ward's algorithm for obtaining clusters (Sneath & Sokal, 1973).

The data were firstly tested for normality (Shapiro-Wilk test) and homoscedasticity (Cochran test) and subjected to ANOVA. The data obtained in each test were analyzed separately by means of analysis of variance and the means of the treatments were compared by the Scott Knott test at 5% probability, through AgroEstat software. Principal component analysis was processed in the STATISTICA software.

Results and discussion

The initial moisture content of the seed lots of *P. sativum* subsp. *Arvense* showed low variation: from 14.0 to 14.6% before aging and from 17.1 to 18.8% after aging at 42 °C for 48 hours (Table 1). These values are within the range of two percentage points, which confer reliability to the test results (Marcos Filho, 2015).

Table 1 - Degree of moisture previous (DMP) and after (DMA) the seeds were aged, thousand seed weight (TSW), field seedling emergence (FSE), first count (FFEC) and index of emergence speed of field seedlings (ESI), germination (G), first germination count (FGC), electrical conductivity (EC) and accelerated aging (AA) of 10 lots of seeds of *Pisum sativum* subsp. *arvense*

Lots	DMP	DMA	TSW	FSE	FFEC	ESI	G	FGC	EC	AA
1	14.1	17.1	149.04 a	94 a	64 c	15.84 a	97 a	47 c	18.18 b	81 a
2	14.3	17.5	146.50 b	93 a	65 c	14.79 b	100 a	42 d	18.67 b	80 a
3	14.0	17.8	150.24 a	86 b	61 c	14.74 b	98 a	56 b	12.94 a	86 a
4	14.1	17.5	144.40 c	87 b	71 b	15.80 a	94 b	66 a	18.15 b	72 b
5	14.4	18.8	150.04 a	89 b	78 a	16.04 a	92 b	32 e	16.79 b	73 b
6	14.1	18.1	107.44 f	88 b	55 d	14.85 b	95 b	42 d	21.24 c	75 b
7	14.6	18.7	96.03 g	86 b	68 c	14.94 b	85 c	45 c	23.87 c	73 b
8	14.0	17.1	94.04 h	80 c	54 d	13.17 c	76 d	44 c	28.41 d	69 b
9	14.2	18.7	136.66 d	82 c	63 c	13.87 b	84 c	31 e	28.02 d	64 c
10	14.2	18.5	133.13 e	68 d	38 e	11.06 d	75 d	38 d	38.71 e	47 d
CV(%)			0.95	4.03	5.67	4.48	2.80	7.33	15.00	8.93

Means followed by different letters in the column differ from each other by the Scott-Knott test (p < 0.05).

Univariate statistics, by the Scott-Knott test, ranked the lots at quality levels. However, the selection of superior lots and efficient variables for this purpose is complicated due to their numbers and the need to know all the information provided by the set of variables; thus, cluster analysis becomes more feasible. This fact was emphasized in soybean (Barbosa et al., 2013) and brachiaria (Silva et al., 2017).

It is worth noting that the tests of interest in the evaluation of lots in seed analysis laboratories, besides ranking the lots at vigor levels, must present results correlated with seedling emergence in the field (Marcos Filho, 2015). Therefore, seedling emergence in the field was used as reference for the selection of vigor tests (Melo et al., 2017; Silva et al., 2017). The performance of the lots was classified in descending

order of vigor, from 1 to 10, grouping them into four vigor classes: high (lots 1 and 2), medium high (lots 3 to 7), medium low (lots 8 and 9), and low (lot 10). This criterion was used because one of the purposes of the vigor tests is to evaluate the physiological quality of the seeds attempting to differentiate lots regarding the vigor level and to predict seedling emergence in the field (Marcos Filho, 2015; Oliveira et al., 2014; Melo et al., 2017; Silva et al., 2017).

In this context, only seedling dry weight and first germination count did not correlate with the reference test (Table 2). It should be noted that to distinguish the more vigorous lots from the less vigorous ones it is necessary to apply a set of vigor tests, adopting at least three of them (Marcos Filho, 2015).

Table 2 - Pearson correlation coefficient (r) between the results of laboratory tests and the emergence of seedlings in the field for 10 seed lots of *Pisum sativum* subsp. *arvense*

Laboratory tests x Emergency of seedlings in the field	r
Thousand-seed weight	0.25 ^{ns}
Germination	0.77 ^{**}
First germination count	0.23 ^{ns}
First count of seedling emergence in the field	0.63 ^{**}
Emergency speed indice in the field	0.88 ^{**}
Electrical conductivity	-0.74 ^{**}
Accelerated aging	0.70 ^{**}

^{ns} e ^{**}: non-significant and significant by F test (p < 0.01), respectively.

As Pearson's correlation, multivariate analysis is a promising statistical tool for the evaluation of vigor tests, as verified in the literature (Barbosa et al., 2013; Silva et al., 2017). It is noteworthy that Pearson's correlation is verified between two variables, while multivariate analysis allowed a broader approach considering several variables.

To explain data variability in the principal components analysis of the 10 forage pea lots, two components with eigenvalues of 5.40 and 0.96 were required;

with total variance of 67.50 and 12.04%. The sum of these values totaled 73.47% of cumulative variance (Table 3). The analysis was adequate because the values of the principal components met the recommendation of Rencher & Christensen (2012), in which approximately 70% of the total variance should be explained by the principal components. That is, a relatively small number of components were extracted (CP1 and CP2) with the ability to explain the greater variability in the original data.

Table 3 - Correlation between each main component and evaluation of the physiological performance of seeds of 10 lots of *Pisum sativum* subsp. *Arvense*.

Variables	Main componente	
	1	2
Thousand-seed weight	0.45	-0.22
Germination	0.92	0.03
First germination count	0.41	0.87
Emergency of seedlings in the field	0.94	-0.14
First count of seedling emergence in the field	0.81	-0.32
Emergency speed indice in the field	0.94	-0.11
Electrical conductivity	-0.97	-0.07
Accelerated aging	0.91	0.15
Eigenvalues	5.40	0.96
Total variance (%)	67.50	12.04
Accumulated variance (%)	79.54	

In determining the number of principal components, those with eigenvalues greater than one are considered (Hongyu et al., 2015). In the present research, the first two components were considered, and the second presented a value close to one. However, the first two principal components effectively summarize the total sample variance and can be used to study the dataset. Principal component 1 is usually the most important, being that of greater variance and which explains the maximum data variability; while the second is the one with the second largest variance (Hongyu et al., 2015).

Similarly, studies on vigor tests for soybean seeds (Barbosa et al., 2013) and *piatã* grass (Silva et al., 2017) also found that two principal components were sufficient to explain, respectively, 65.86 and 74.23% of the variance in the discrimination of variables. In contrast, in similar work with rice, Lorentz &

Nunes (2013) verified the need for three principal components, which explained 80.42% of the characteristics.

For each of the principal components, all correlation values equal to or greater than 0.6 were considered relevant and with discriminatory power (Lorentz & Nunes, 2013; Silva et al., 2017). Thus, among the variables, only thousand seed weight had no discriminatory power in the two principal components.

According to Hongyu et al. (2015), the discriminatory power of the variables in each principal component is measured by correlation. Therefore, based on the values in Table 3, it can be inferred that the variables germination (0.92), seedling emergence in the field (0.94), first count of seedling emergence in the field (0.81), seedling emergence speed index in the field (0.94), electrical conductivity (-0.97), and accelerated aging (0.91) were correlated with principal com-

ponent 1. In turn, only the first germination count (0.87) was correlated with principal component 2.

The variable electrical conductivity had values inversely proportional to the others in principal component 1. This can be verified by the negative and positive values of these variables, respectively. Therefore, the higher the electrical conductivity of the solution, the lower the seed quality of this lot regarding germination and vigor tests. This fact is widely validated in the literature (Machado et al., 2011; Marcos Filho, 2015).

The set of eight variables in the evaluation of the quality of forage pea seeds allowed the principal components analysis to synthesize the maximum of the original information into two latent variables called principal components, which enabled their location in two-dimensional figures (ordering of accesses of the principal components), as shown in Figure 1. According to Hair et al. (2009), these are linear combinations of the original variables created with the two largest eigenvalues of the covariance matrix.

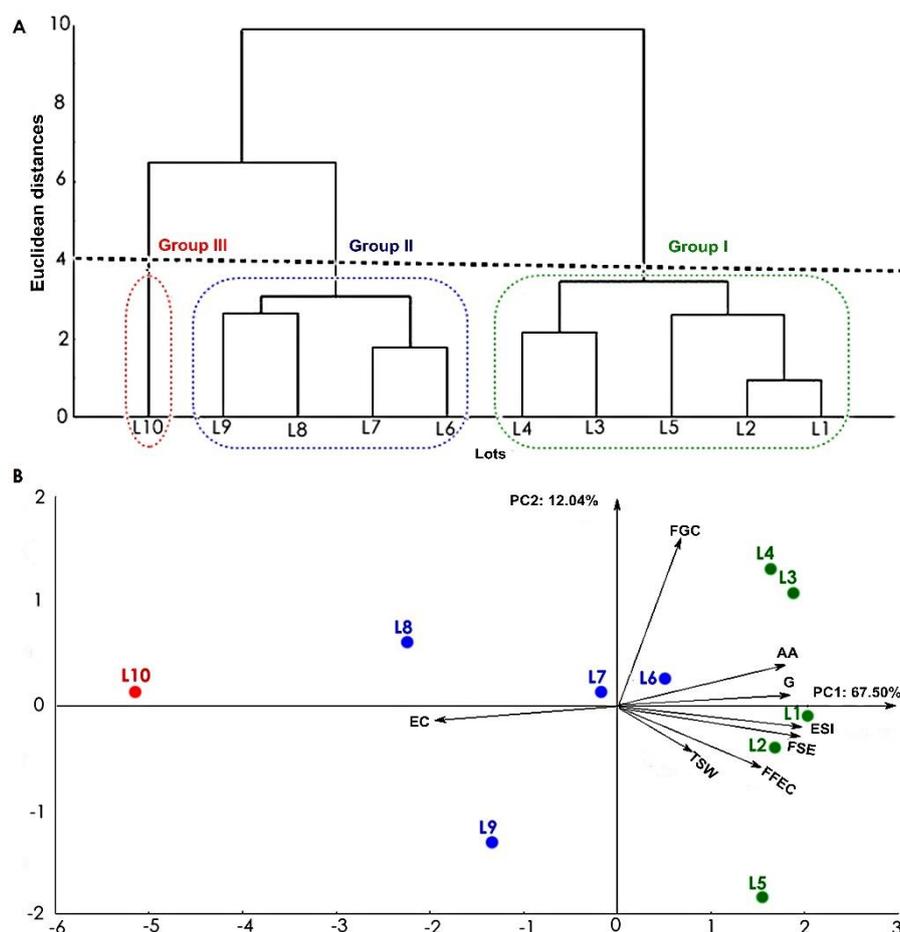


Figure 1 - Dendrogram resulting from hierarchical grouping analysis with the formation of groups (A), circle of eigenvectors and dispersion plane (B) obtained by analysis of principal components according to Thousand-seed weight (TSW), germination (G), first seed germination count (FGC), field seedling emergence (FSE), first field seedling emergence count (FFEC), field seedling emergence index (ESI), electrical conductivity (EC) and accelerated aging (AA) of seeds of 10 lots of *Pisum sativum* subsp. *arvense*

The methods of multivariate data analysis allow a global study of these variables, evidencing the connections, similarities, or differences between them, with the least possible loss of information. Due to the significant variation in the Euclidean distance between the seed lots, it was possible to divide the original set into three groups, adopting the Euclidean distance 4 (Figure 1A) (Hair et al., 2009).

These groups were well distinguished in the scatter plot (Figure 1B). Group I was composed of seeds from lots 1 to 5, which presented higher values of germination and vigor, indicated by the eigenvectors

of these variables closer to these production sites, and lower electrical conductivity, with eigenvector in the quadrant opposite the lots. This fact was validated by the means comparison test (Tables 1 and 2).

Group II was formed by lots 6 to 9, with intermediate quality according to the values of germination and vigor. In turn, group III was composed of lot 10, classified as lower quality, i.e., with lower values of germination and vigor, mainly due to the high electrical conductivity in the solution of the evaluated seeds (Figure 1).

High correlations were observed among the variables in the dispersion plane: germination, first count of seedling emergence in the field, seedling emergence speed index in the field, and accelerated aging, since the arrows indicative of these parameters formed acute angles between them. In a different way, the variable electrical conductivity showed high correlation, although negative, with seedling emergence in the field, since the arrows formed angles near 180 degrees for this parameter.

Although presenting correlation with seedling emergence in the field, the variable thousand seed weight is not considered efficient to rank forage pea seed lots since smaller eigenvectors were observed, indicating a smaller variation among the lots and, therefore, less power to differentiate them.

First germination count was lowly correlated with seedling emergence in the field, since the arrows formed angles near 90 degrees for this parameter.

Conclusions

The multivariate analysis of principal components is efficient in discriminating seed vigor and germination tests in *Pisum sativum* subsp. *arvense*, which identify lots of superior performance in the field.

References

- Barbosa RM, Vieira BGTL, Ferraudo AS, Corá JE, Vieira RD (2013) Discrimination of soybean seed lots by multivariate exploratory techniques. *Journal of Seed Science*. 35(3): 302-310. doi: 10.1590/S2317-15372013000300005
- Brasil. 2009. Regras para análise de sementes. MAPA-ACS, Brasília, Brasil. 395p.
- Cargnelutti Filho A, Alves BM, Burin C, Kleinpaul JA, Neu IMM, Silveira DL, Simões FM, Spanholi R, Medeiros LB (2015) Tamanho de parcela e número de repetições em ervilha forrageira. *Ciência Rural* 45(7): 1174-1182. doi: 10.1590/0103-8478cr20141043
- Fávero LP, Belfiore P (2015) Análise de Dados: Técnicas Multivariadas Exploratórias com SPSS e STATA. Elsevier Editora, Rio de Janeiro, Brasil. 368p.
- Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2009) Análise Multivariada de dados. Porto Alegre: Editora Bookman, 688p.
- Hongyu K, Sandanielo VLM, Oliveira Junior GJ (2015) Análise de componentes principais: resumo teórico, aplicação e interpretação. *Engineering and Science* 5(1): 83-90. doi: 10.18607/ES201653398
- Lorentz LH, Nunes UR (2013) Relações entre medidas de qualidade de lotes de sementes de arroz. *Revista Ciência Agrônômica* 44(4): 798-804.
- Machado CG, Silva GZ, Oliveira SSC, Martins CC (2019) Accelerated aging test for forage pea seeds. *Semina: Ciências Agrárias* 40(5): 1819-1828. doi: 10.5433/1679-0359.2019v40n5p1819
- Machado CG, Martins CC, Santana DG, Cruz SCS, Oliveira SSC (2011) Adequação do teste de condutividade elétrica para sementes de *Pisum sativum* subsp. *arvense*. *Ciência Rural* 41(6): 988-995. doi: 10.1590/S0103-84782011005000062
- Maguire JD (1962) Speed of germination: Aid in selection and evaluation for seedling emergence and vigour. *Crop Science* 2(2): 176-177.
- Marcos Filho J (2015) Seed vigor testing: an overview of the past, present and future perspective. *Scientia Agricola* 72(4): 363-374. doi: 10.1590/0103-9016-2015-0007
- Melo PAFR, Jeromini TS, Affonso CE, Alves EU, Martins CC (2017) Vigor tests in assessing the quality of signal grass seeds. *Semina: Ciências Agrárias* 38(6): 3491-3498. doi: 10.5433/1679-0359.2017v38n6p3491
- Oliveira SSC, Martins CC, Cruz SJS, Silva CJ (2014) Seleção de progênies de nabo-forrageiro para germinação sob altas temperaturas. *Ciência Rural* 44(2): 217-222. doi: 10.1590/S0103-84782014000200004
- Pinto AH, Binotti FFS, Souza HM, Batista TB, Gouveia GCC (2016) Teste de condutividade elétrica para diferenciação dos níveis de deterioração de sementes de forrageiras. *Revista de Agricultura Neotropical* 3(2): 9-15. doi: 10.32404/reatn.v3i2.1070
- Rencher AC, Christensen WF (2012) *Methods of Multivariate Analysis*. John Wiley Professio, New York, USA. 758p.
- Santos A, Ceccon G, Davide LMC, Neto Neto AL, Silva JF, Alvarez RCF (2016) Integrated multivariate analysis to identify superior cowpea genotypes. *Comunicata Scientiae* 7(2): 223-231. doi: 10.14295/cs.v7i2.643
- Silva GZ, Martins CC, Cruz JO, Jeromini TS, Bruno RDLA (2017) Evaluation the physiological quality of *Brachiaria brizantha* cv. BRS 'Piatã' seeds. *Bioscience Journal* 33(3): 572-580. doi: 10.14393/BJ-v33n3-36519
- Sneath PHA, Sokal RR (1973) *Numerical taxonomy: the principles and practice of numerical classification*. Freeman, San Francisco, USA. 573p.
- Souza IGB, Souza VAB, Silva KJD, Lima PSC (2016) Multivariate analysis of 'bacuri' reproductive and vegetative morphology. *Comunicata Scientiae* 7(2): 232-240. doi: 10.14295/CS.v7i2.779