Article

The use of a multivariate statistical procedure in analysing the germination process of two bean cultivars, compared with a univariate approach

Andre Bianconi[1] , Jose S. Govone[2], Bryan F. J. Manly[3], Michael J. Watts[4]

[1]Department of Zoology, Biosciences Institute – Unesp – São Paulo State University, Avenida 24-A, 1515, Bela Vista, Rio Claro-SP, Brazil

[2]DEMAC, IGCE – Unesp – Rio Claro-SP, Brazil

[3]Western EcoSystems Tech. Inc., 2003, Central Avenue Cheyenne, Wyoming, 82001, USA

[4]School of Earth and Environmental Sciences, The University of Adelaide, Australia

E-email: drebianconi@yahoo.com.br

Received 13 May 2011; *Accepted 1 June 2011*; *Published online 28 August 2011*

IAEES

Abstract

Several studies on plant physiology are aimed at describing or assessing seed germination processes under laboratory conditions. With respect to seed germination of crop species, some statistical complexities have been discussed, but they have not been developed much in practice. That is, such discussions are not as common as in other areas of plant biology. Additionally, the current literature that is concerned directly with the application of statistics in seed germination indicates that simple and well-known statistical procedures still merit further consideration. Regarding the use of multivariate statistical methods in agriculture, several field studies have used such procedures as a means of clarifying some underlying ecological principles that govern crop production. Nonetheless, multivariate tests have not been widely employed in germination experiments. Therefore, in the present study a simple multivariate statistical procedure (Hotelling's T^2 statistic) was utilised in order to compare two common bean cultivars, using germination indices as variables. The outcome derived from the multivariate approach was compared with that obtained from the utilisation of the univariate *t* test. The simultaneous application of both methods (that is, the classical univariate *t* test and the multivariate T^2 test) showed that the outcomes may well depend on the approach utilised.

Key words germination rate; Hotelling's T^2; synchronisation; *t* test.

1 Introduction

The common bean (*Phaseolus vulgaris* L.) is recognised worldwide both as an important subsistence crop and as a profitable agricultural product. In Brazil, for example, common bean cultivars are utilised by most of the population as an accessible source of protein. Furthermore, extensive agronomic research into the life cycle of this species is regarded as essential, because of the great usefulness of this crop as a feeding resource (Ramalho and Abreu, 2006). On the whole, the current production and distribution of beans is directly influenced by several socioeconomic and environmental factors that may be fundamentally interrelated. To a large extent, such factors are responsible for the interest in studying every developmental stage of *P. vulgaris* in order to enhance and forecast its yield. In this integrated context, seed germination represents an important developmental phase that should be described in a detailed manner.

It is important to note that several studies are aimed at describing or assessing seed germination processes under laboratory conditions (Labouriau and Agudo, 1987; Moreau-Valancogne et al., 2007) or partially controlled conditions (e.g., Mavi et al., 2010). The main drawback to germination studies conducted under controlled conditions is that they are frequently confined to artificial factors, and such a fact precludes general assertions concerning the potentialities of specific crops. Moreover, crop yield depends on factors that are not restricted to the germination phase (Graham and Ranalli, 1997), and the variable behaviour of field conditions may not be satisfactorily simulated under controlled conditions.

The extrapolation of results derived from controlled experiments (especially in the case of cash crops) suffers from limitations that might not be suitably overcome. Nonetheless, germination studies conducted under laboratory conditions possess the advantage of permitting researchers to obtain larger sample sizes than field experiments. In view of the fact that adequate sampling is essential for every statistical procedure, laboratory experiments provide seed physiologists with the possibility of deploying experimental designs that would not be feasible under natural conditions.

In regard to seed germination parameters of weeds and crops, some statistical complexities and implications have been discussed, but they have not been developed much in practice (e.g., Santana and Ranal, 2006; Gardarin et al., 2011). That is, such discussions are not as common as in other areas of plant biology. With respect to the utilisation of multivariate statistical methods in agriculture, several field studies have used such statistical procedures as a means of clarifying some underlying ecological principles that govern crop production. Interestingly, multivariate tests have been feasibly applied to novel approaches such as agroecology and organic farming. For instance, Drinkwater et al. (1995) used classical multivariate statistical approaches in order to identify relationships between agronomic and community level properties of organic and conventional systems, and their assessment was improved with the outcomes derived from the implementation of principal component analysis, canonical discriminant analysis, and non-hierarchical cluster analysis.

Furthermore, principal component analysis was utilised by Monokrousos et al. (2008) with the aim of comparing organic cultivation with conventional agricultural practices, using soil biochemical variables. This multivariate statistical analysis permitted the authors to assert that the changes in soil biochemical variables with years of organic management were more complex than they had initially expected.

Nonetheless, multivariate tests have not been widely employed in germination experiments, especially studies conducted under laboratory conditions, in which the evaluation and measurement of several variables could well be easier than in most of the current field experiments. Therefore, in the present study a simple multivariate statistical procedure (T^2 statistic) was utilised in order to compare two common bean cultivars, using well-known germination indices as variables (germination rate and synchronisation). Additionally, the outcome derived from the multivariate approach was compared with the result obtained through the utilisation of the classical univariate *t* test. The option of presenting a simple and controlled experiment was chosen because it is hoped that such analyses, albeit under artificial conditions, will serve as a basis for deploying more complex designs in a feasible manner.

2 Materials and Methods

2.1 Seed material

'Iapar 81' and 'IPR Uirapuru' were the two bean cultivars analysed, and each of them constitutes a treatment in the present work. Both cultivars were obtained from Instituto Agronomico do Parana (Londrina, Parana state, Brazil) and harvested in November 2006. The viability of the utilised seed lots was assessed by means of the tetrazolium test ($C_{19}H_{15}N_4Cl$, 0.3%, data not shown) prior to the beginning of the experiment (see Sawma and Mohler, 2002, for a discussion about seed viability testing). Eighteen replicates were utilised in every treatment, and each replicate consisted of a Petri dish (9cm in diameter, 12mm in depth) containing three layers of filter paper (moistened with distilled water) on which 20 seeds were set out. Each Petri dish was put into a transparent plastic germination box and kept at a temperature of 25±2ºC under constant white light. A seed was characterised as germinated after the protusion of its radicle. The germination counts were performed at 24-hour intervals, and the germinated seeds were progressively removed from the dishes.

2.2 Germination indices

Germination rate and synchronisation were the two variables analysed in this paper. The germination rate ($\overline{V}$) is usually defined as the inverse of the mean time of germination (mean time^{-1}, measured in days^{-1} in this study). That is,

$$\text{mean time} = \left[\left(\sum_{i=1}^{n} n_i T_i \right) \bigg/ \left(\sum_{i=1}^{n} n_i \right) \right] \qquad (1)$$

in which T_i is the elapsed time from the start of the experiment to the n_{th} observation (days or hours); n_i is the number of seeds germinated at time i (the number of seeds that germinated within a specific period of observation) (Ranal and Santana, 2006). Mavi et al. (2010), for example, utilised a mean time germination index similar to the mean time deployed in this paper.

The uniformity or synchronisation of germination ($\bar{E}$) is defined as

$$\overline{\boldsymbol{E}} = -\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \boldsymbol{f_i} \log_2(\boldsymbol{f_i}) \qquad (2)$$

in which f_i represents the relative frequency of germination (that is, the number of germinated seeds within a specific time interval (i) divided by the total number of germinated seeds over the whole experiment); and log_2 is the base-2 logarithm of each fi. Low synchronisation values indicate more synchronised germination than high values. 'Bit' represents the unit of synchronisation if $\log_2$ is utilised (see Ranal and Santana, 2006, for additional details).

The attempts to incorporate two distinct characteristics of the germination process (germination rate and synchronisation, for example) in the same mathematical expression have not been entirely satisfactory (Bewley and Black, 1994), and such a fact may be the reason behind the deployment of germination indices that describe one single aspect of the germination process. For instance, 'germination rate' describes only the germination speed, and 'synchronisation' describes whether or not the germination occurs in a synchronised manner. However, the germination rate and synchronisation have been widely utilised in seed physiology. There has been some debate over the usefulness of these indices as effective germination measures; nonetheless, such indices are deemed to be important variables in seed germination (Labouriau and Agudo, 1987; Ranal and Santana, 2006).

Moreover, the comprehension of the germination rate and synchronisation plays an important role in forecasting productivity levels of crops, because the dynamics of every developmental stage should be carefully analysed with the aim of obtaining a high crop yield in the long term (Gardarin et al., 2011). Both indices are of equal biological value (Labouriau and Agudo, 1987; Larcher, 1994; Ranal and Santana, 2006). Therefore, they were both utilised in the present description of how a simple multivariate test could be deployed in comparing two bean cultivars with respect to their germination phase.

2.3 Statistical procedure

Hotelling's T^2 statistic was used in order to verify whether or not the difference between 'Iapar 81' and 'IPR Uirapuru' was significant with respect to mean values of two variables (i.e. germination indices). The T^2 statistic is a multivariate generalisation of the well-known univariate *t* test and is clearly and thoroughly described in a large number of statistical texts (e.g., Mardia et al., 1979; Manly, 2005). The significance of T^2, or the lack thereof, may be easily determined if we bear in mind that, in the null hypothesis case of equal population means, the transformed statistic $F = [(n_1+n_2-p-1)\ T^2\ ] / [(n_1+n_2-2)\ p]$ follows a statistical F distribution with p and (n_1+n_2-p-1) degrees of freedom, in which n_i represents the sample size of each treatment, and p represents the number of variables (Manly, 2005). This transformation is not essential for employing this test, but it permits researchers to obtain a *P*-value using readily available spreadsheet programs (see Mardia et al., 1979, and Manly, 2005, for additional details concerning the computation of Hotelling's T^2).

The outcome derived from Hotelling's T^2 was compared with the results obtained from the well-known two-sample *t* test (Student's *t* test) (see Zar, 1999, for details). The univariate *t* test was utilised in order to compare the two cultivars in relation to each variable separately. All statistical analyses were conducted at a 5% significance level.

The robustness of Hotelling's T^2 is discussed in some papers (e.g., Mardia, 1975, and Kariya, 1981). To some extent, if each one of the variables being analysed appears to be normally distributed, then the joint distribution of such variables is usually regarded as multivariate normal. Furthermore, it is important to note that, in general, multivariate tests seem to be less robust than univariate tests in terms of the normal distribution assumption (Manly, 2005). However, this does not mean that several univariate tests should substitute for a single multivariate analysis on the basis that univariate analyses seem to be more robust with respect to the assumption of normality. In fact, the choice between univariate and multivariate methods should be based on the effectiveness of such approaches. For example, most ecological phenomena could be understood only by means of assessing the joint influence of the variables being analysed (Zhang, 2011), which could require using an appropriate multivariate technique.

3 Results

Regarding the multivariate T^2 test, neither cultivar was statistically different (T^2 = 4.92, P=0.11). That is, if we analyse the simultaneous influence of the two variables (germination rate and synchronisation) on the germination response of the beans, the two cultivars will be deemed to be similar. On the other hand, if we were to deploy the univariate *t* test, the outcome would be different from that derived from the multivariate approach, because the cultivar 'IPR Uirapuru' exhibits a slightly more synchronised (smaller mean value) germination than 'Iapar 81' (t=2.11, P=0.04) (Table 1). Thus, with respect to the two germination indices, both cultivars present distinct responses from a univariate point of view. Conversely, from a multivariate standpoint, neither cultivar was statistically different.

Table 1- Two *Phaseolus vulgaris* cultivars ('Iapar 81' and 'IPR Uirapuru') and their mean and standard error values for the variables germination rate (X_1) and synchronisation (X_2).

	Iapar 81		IPR Uirapuru	
	X_1	X_2	X_1	X_2
Mean*	0.27^a	1.49^b	0.31^a	1.24^c
Standard error	0.02	0.09	0.01	0.08

*Different letters represent cultivars that are statistically different in relation to each variable separately, using the univariate *t* test. Each sample size was equal to 18, and 5% was the significance level.

4 Discussion

Santana and Ranal (2006), for example, discussed several implications of using the simple linear correlation in germination experiments. Such a test has been broadly used and described in the literature. Nonetheless, it was reconsidered, and the authors took into account the great interference of the sample size in the significance of the correlation analysis. This type of evaluation of a trivial statistical test implies that the implementation of simple experimental designs in germination experiments still deserves serious consideration. Furthermore, the employment of complex statistical techniques, such as most of the multivariate statistical methods, may well be considered a novel approach in comparative germination analysis.

In the current study, Hotelling's T^2 test assessed the evidence from all of the data together, but each *t* test was only looking at half of the data. Thus, it should not be surprising that the results were not the same. Regarding the T^2 test, the significant difference on one variable was offset by the means being very similar on the other variable (Table 1). From an ecological point of view, both variables (germination rate and synchronisation) are deemed to be equally important (Larcher, 1994). Therefore, the multivariate approach may well be deemed to be more plausible for a general comparison between the two cultivars than the univariate analysis, even though the univariate method could not be considered as an unsuitable or limited option in the strictest sense.

In the univariate case, it is important to highlight that the cultivars should be compared between themselves in relation to each variable separately. Moreover, it should be noted that the T^2 statistic allows for correlation between variables, but individual *t* tests do not. In other words, a multivariate test is not a simple combination of univariate tests (Manly, 2005). Therefore, the T^2 statistic allows the researcher to analyse the cultivars in relation to the germination rate and synchronisation in a simultaneous manner.

Regarding the reasonable implementation of hypothesis tests, it is important to emphasise that small sample sizes should not be used. There are relevant studies on the germination of crop seeds in which the sample sizes vary from four to six replicates of each treatment. Such studies (e.g., Gummerson, 1986; Alvarado and Bradford, 2002; Mavi et al., 2010) are not especially aimed at assessing the power-efficiency of hypothesis tests, and thus such small sample sizes might not be deemed to be restrictive factors, because biological aspects of germination processes could be more important than possible statistical implications from a strict biological point of view. Nonetheless, 4-6 replicates per treatment may be considered insufficient for analysing germination processes (irrespective of the number of seeds per replicate), and such small sample sizes could not provide seed physiologists with a reliable statistic, even if nonparametric hypothesis tests were deployed. Therefore, larger sample sizes are highly desirable in order to conduct feasible and accurate germination experiments (Gardarin et al., 2011).

Simple physiological studies of the germination of several types of seeds are frequently found in the literature. In relation to *P. vulgaris*, descriptive and analytical studies of the influence of temperature and water potential on seed germination and hypocotyl elongation under laboratory conditions are relatively common (Moreau-Valancogne et al., 2007). Furthermore, the investigation of the secondary structure of different proteins (Carbonaro et al., 2008) and the examination of the phenolic profiles of dry bean cultivars (Lin et al., 2008) represent biochemical approaches to the basic characterisation of beans.

Similarly, other important crop species have been analysed with the aim of assessing the effect of temperature and pH value of stress medium on hydroxyl radical inhibition (Li et al., 2008). Mavi et al. (2010), for instance, evaluated mean germination time as a feasible test in order to rank the relative emergence of seeds under stress conditions, and their work indicated that differences in mean germination times between seed lots may determine the emergence performance regarding three cucurbit crops. These examples could indicate that crop seeds and their germination processes are frequently analysed in a detailed manner, with the appropriate application of multivariate methods when necessary.

However, the germination of seeds has not been studied by means of multivariate approaches. In addition, if multivariate statistical tests were substituted for univariate statistical methods (whenever necessary) in the basic characterisation of bean cultivars, this substitution would provide researchers with more realistic outcomes. From an agroecological point of view, the combined response of germination indices to different treatments may well be more important than their separate responses, because several variables determine the features of a specific seed simultaneously, and this intrinsically multivariate phenomenon will be dealt with in an unrealistic univariate manner if a univariate test is deployed in this context.

Both cultivars (‘Iapar 81’ and ‘IPR Uirapuru’) are usually described as moderately tolerant of water stress and high temperatures (IAPAR, 2011). However, such general characteristics were derived from field experiments conducted under the direct influence of several variables, and they were not based on detailed germination experiments. Therefore, the present study, albeit under artificial conditions, may be useful to broaden the current knowledge of the developmental stages of bean cultivars.

5 Concluding Remarks

The current work described a simple experiment, and other practical situations may require the utilisation of other variables, as well as more sophisticated experimental designs. Apart from their biological and ecological importance, the two variables used in this analysis were chosen for ease of presentation, because the main aim of this research was to introduce a simple multivariate statistical method into the literature on seed germination. In addition, this study was aimed at encouraging seed physiologists to deploy multivariate methods in characterising crop seeds and their germination processes, because multivariate tests may be more realistic than a univariate approach in several practical situations.

The simultaneous application of both methods (that is, the univariate t test and the multivariate T^2 test) presented in this paper showed that the outcomes may well depend on the approach utilised. Owing to the importance of *P. vulgaris* as a feeding resource (Ramalho and Abreu, 2006), the more effective the statistical method, the less scope there is for an unrealistic description, and the same is true of other important crops.

Acknowledgements

Fapesp (Fundação de Amparo à Pesquisa do Estado de São Paulo) provided the first author with financial support.

References

Alvarado V, Bradford KJ. 2002. A hydrothermal time model explains the cardinal temperatures for seed germination. Plant, Cell and Environment, 25: 1061-1069

Bewley JD, Black M. 1994. Seeds: physiology of development and germination (2nd ed). Plenum Press, New York, USA

Carbonaro M, Maselli P, Dore P, et al. 2008. Application of Fourier transform infrared spectroscopy to legume seed flour analysis. Food Chemistry, 108: 361-368

Drinkwater LE, Letourneau DK, Workneh F, et al. 1995. Fundamental differences between conventional and organic tomato agroecosystems in California. Ecological Applications, 5: 1098-1112

Gardarin A, Dürr C, Colbach N. 2011. Prediction of germination rates of weed species: Relationships between germination speed parameters and species traits. Ecological Modelling, 222: 626-636

Graham PH, Ranalli P. 1997. Common bean (*Phaseolus vulgaris* L.). Field Crops Research**,** 53: 131-146

Gummerson RJ. 1986. The effect of constant temperatures and osmotic potentials on the germination of sugar beet. Journal of Experimental Botany, 37: 729-741

IAPAR. 2011. Características das cultivares de feijão com sementes no mercado [Characteristics of bean cultivars that possess seeds available on the market]. Instituto Agronômico do Paraná, Londrina, Paraná state, Brazil. Information collected on 23-03-2011. http://www.iapar.br/modules/conteudo/conteudo.php?conteudo=67

Kariya T. 1981. A robustness property of Hotelling's T^2-test. The Annals of Statistics, 9: 211-214

Labouriau LG, Agudo M. 1987. On the physiology of seed germination in *Salvia hispanica* L. I. Temperature effects. Anais da Academia Brasileira de Ciências, 59: 37-56

Larcher W. 1994. Ökophysiologie der Pflanzen: Leben, Leistung und Streßbewältigung der Pflanzen in ihrer Umwelt. Eugen Ulmer GmbH, Stuttgart, Germany

Li J, Chen Z, Guan X, et al. 2008. Optimization of germination conditions to enhance hydroxyl radical inhibition by water-soluble protein from stress millet. Journal of Cereal Science, 48: 619-624

Lin LZ, Harnly JM, Pastor-Corrales MS, et al. 2008. The polyphenolic profiles of common bean (*Phaseolus vulgaris* L.). Food Chemistry, 107: 399-410

Manly BFJ. 2005. Multivariate Statistical Methods: a Primer (3rd ed). Chapman & Hall/CRC, Boca Raton, USA

Mardia KV. 1975. Assessment of multinormality and the robustness of Hotelling's T^2 test. Journal of the Royal Statistical Society (Series C – Applied Statistics), 24: 163-171

Mardia KV, Kent JT, Bibby JM. 1979. Multivariate Analysis. Academic Press, London, UK

Mavi K, Demir I, Matthews S. 2010. Mean germination time estimates the relative emergence of seed lots of three cucurbit crops under stress conditions. Seed Science & Technology, 38: 14-25

Monokrousos N, Papatheodorou EM, Stamou GP. 2008. The response of soil biochemical variables to organic and conventional cultivation of *Asparagus* sp. Soil Biology & Biochemistry, 40: 198-206

Moreau-Valancogne P, Coste F, Dürr C, et al. 2007. Effects of bean seed production conditions on germination and hypocotyl elongation responses to temperature and water potential. In: Seeds: biology, development and ecology (SW Adkins, SE Ashmore, SC Navie eds). CAB International, Cambridge, 333-341

Ramalho MAP, Abreu AFB. 2006. Cultivares [Cultivars]. In: Feijão [Common Bean], second ed (C Vieira, TJ Paula Jr., & A Borém eds). Editora UFV, Viçosa-MG- Brazil, 415-436

Ranal MA, Santana DG. 2006. How and why to measure the germination process? Revista Brasileira de Botânica, 29: 1-11

Santana DG, Ranal MA. 2006. Linear correlation in experimental design models applied to seed germination. Seed Science & Technology, 34: 233-239

Sawma JT, Mohler CL. 2002. Evaluating seed viability by an unimbibed seed crush test in comparison with the tetrazolium test. Weed Technology, 16: 781-786

Zar JH. 1999. Biostatistical Analysis, fourth ed. Prentice Hall, New Jersey, USA

Zhang WJ. 2011. Simulation of arthropod abundance from plant composition. Computational Ecology and Software, 1: 37-48