

*Full Length Research Paper*

# Functionality and efficiency in the analyses of experimental data for Agricultural research

**\*Anderson Inacio, Dilma Ribeiro and Philippe Firmino**

Department of Food Engineering, Federal University of Campina Grande, Campina Grande, Paraíba, Brazil.

Accepted 18 February, 2017

Statistical programs are essential tools for those who deal with scientific research and need to analyze experimental data. In agriculture, there are often uncontrolled factors, which determine the necessity of statistical analyses of the data. The Assistat software version 7.7 is one of these tools and this study aimed to demonstrate its functionality and efficiency in the analyses of experimental data of agricultural research and evaluate its acceptance. In order to exemplify its utilization, data of agricultural experiments were analyzed using the models of analysis of variance for randomized block and factorial experiments. In addition, the regression was used in the analysis of variance for quantitative treatments. It was concluded that the software was used in many papers published in journals and that it is functional and efficient in the analysis of experimental data of agricultural research.

**Key words:** Analysis of variance (ANOVA), statistical software, Tukey's test.

## INTRODUCTION

Tools such as statistical programs are essential for those who deal with agricultural research and need to analyze experimental data. There are good professional programs to meet this demand, such as: SAS, SAEG, STATISTICA, SPSS, XLSTAT and others. Nonetheless, some of these programs have a considerable cost and it makes their acquisition by the students difficult; in addition, SAS and SAEG are packages and it is necessary to write a routine to perform the desired analysis. There are many other free programs available for those who need to analyze

experimental data.

The Assistat software (Silva and Azevedo, 2006) is one more free tool available to meet the demand for data analysis. Although it caters to other areas, it is basically intended for agricultural research, since it contemplates the main models of the analysis of variance (ANOVA) and the like. It has a friendly and easy-to-use interface. Besides the ANOVA models, it has, among others, the principal component analysis (PCA) and the main tests of the non-parametric statistics.

According to Cox and Reid (2000), there is a wide selection of statistical computing packages, and most of these provide the facility for analysis of variance and estimation of treatment contrasts in one form or another.

---

\*Corresponding Author. E-mail: [inacio.dilma@hotmail.com](mailto:inacio.dilma@hotmail.com)

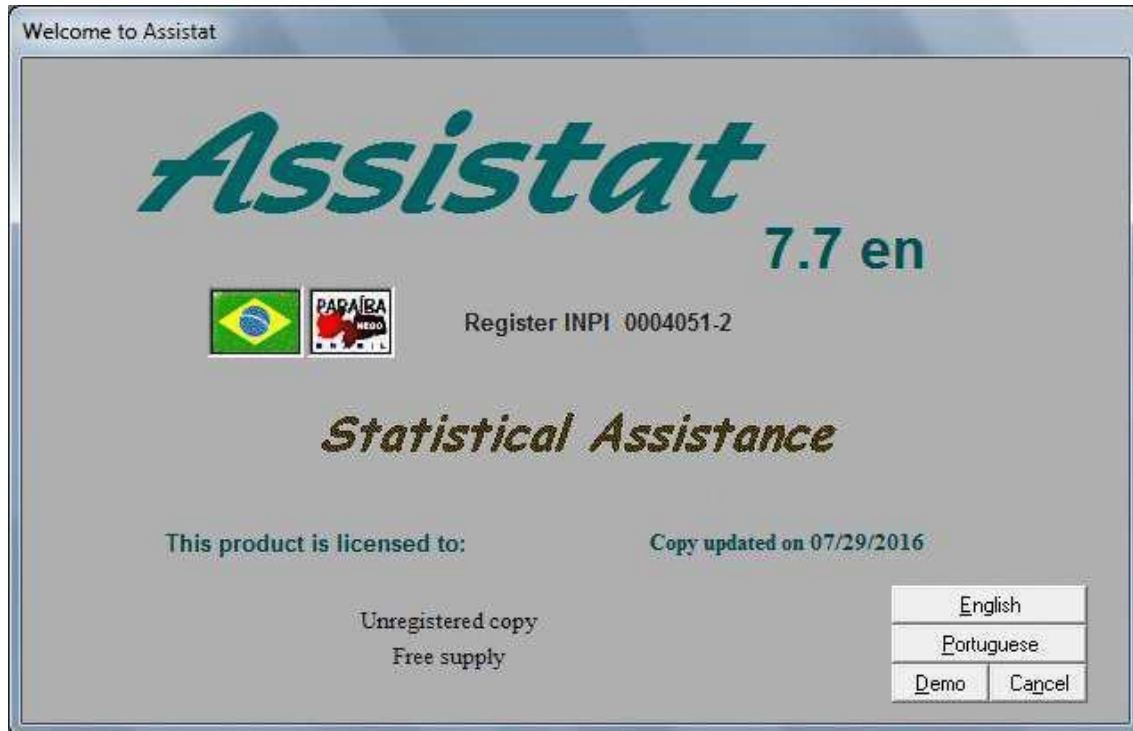


Figure 1. Initial screen.

With small data sets it is often straightforward, and very informative, to compute the contrasts of interest by hand.

In  $2^k$  factorial designs this is easily done using Yates's algorithm. The package GENSTAT is particularly well suited to analysis of complex balanced designs arising in agricultural application. SAS is widely used in North America, partly for its capabilities in handling large databases. GLIM is very well suited to empirical model building by the successive addition or deletion of terms, and for analysis of non-normal models of exponential family form.

The analysis of a complex process requires the identification of target quality attributes that characterize the output of the process and of factors that may be related to those attributes. Once a list of potential factors is identified from subject-matter expertise, the strengths of the associations between those factors and the target attributes need to be quantified. A naïve, one-factor-at-a-time analysis would require many more trials than necessary. Additionally, it would not yield information about whether the relationship between a factor and the target depends on the values of other factors (commonly referred to as interaction effects between factors). As demonstrated in Douglas Montgomery's Design and Analysis of Experiments textbook, principles of statistical theory, linear algebra, and analysis guide the development of efficient experimental designs for factor settings. Once a subset of important factors has been isolated, subsequent experimentation can determine the

settings of those factors that will optimize the target quality attributes. Fortunately, modern software has taken advantage of the advanced theory. This software now facilitates the development of good design and makes solid analysis more accessible to those with a minimal statistical background (Montgomery, 2008).

This study aimed to demonstrate the functionality and efficiency of the Assistat software version 7.7 in the analysis of experimental data of agricultural research and evaluate its acceptance.

## MATERIALS AND METHODS

The Assistat software is an application for Windows and works with TXT files. Data entry can be performed through typing (via keyboard), reading of data in a TXT file and through importation of data from an Excel worksheet. It is freely distributed for physical and legal persons, thus for all of those for whom it can be useful in the analysis of experimental data. Figures 1 to 4 show the main screens of the Assistat. Figure 1 shows the initial screen. Figure 2 shows the options of analysis and tests. Figure 3 is the final menu of an analysis of variance; before the analysis, it is possible to transform the data, verify normality etc. Figure 4 shows the options for the selection of the test of comparison of means; after the selection, the results are presented. On the screen of the results, there is a button called Previous Menu, which allows to go back and select another test of comparison of means.

The utilization of the Assistat software was exemplified using experimental data of the agricultural area from the literature. The Tukey test was used to compare the means, for being the most used among the tests, but it is possible to see that the Assistat

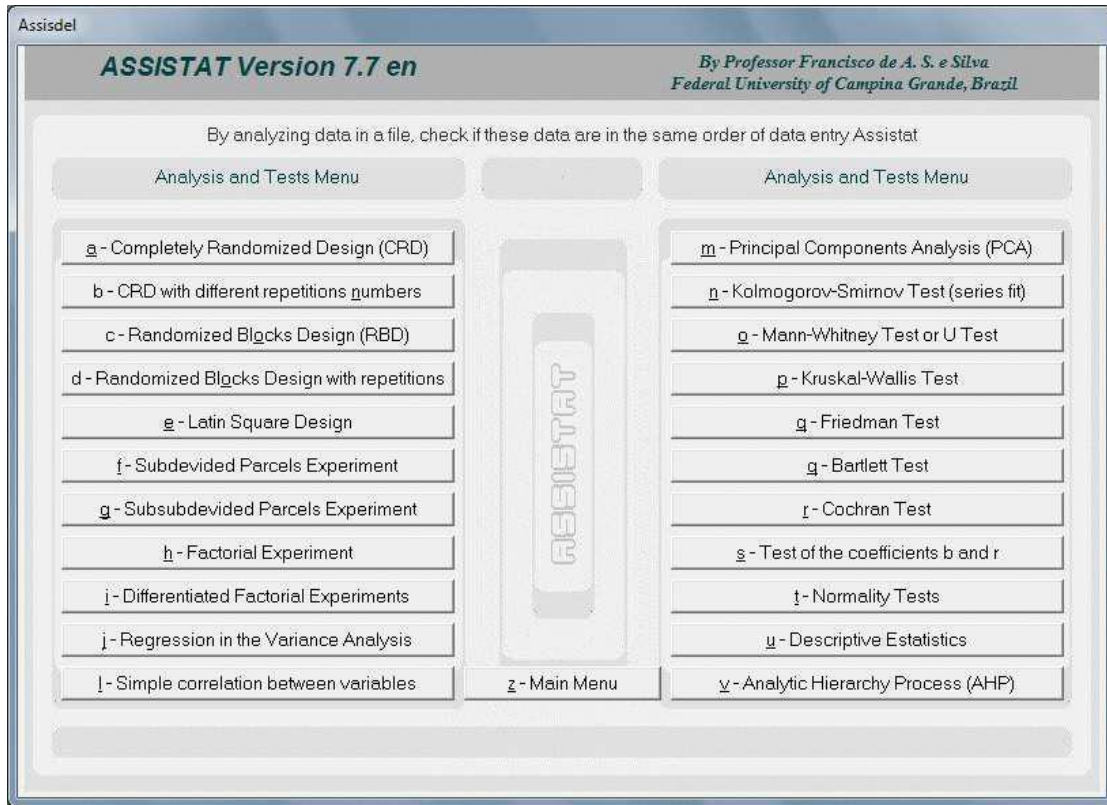


Figure 2. Menu of analyses and tests.

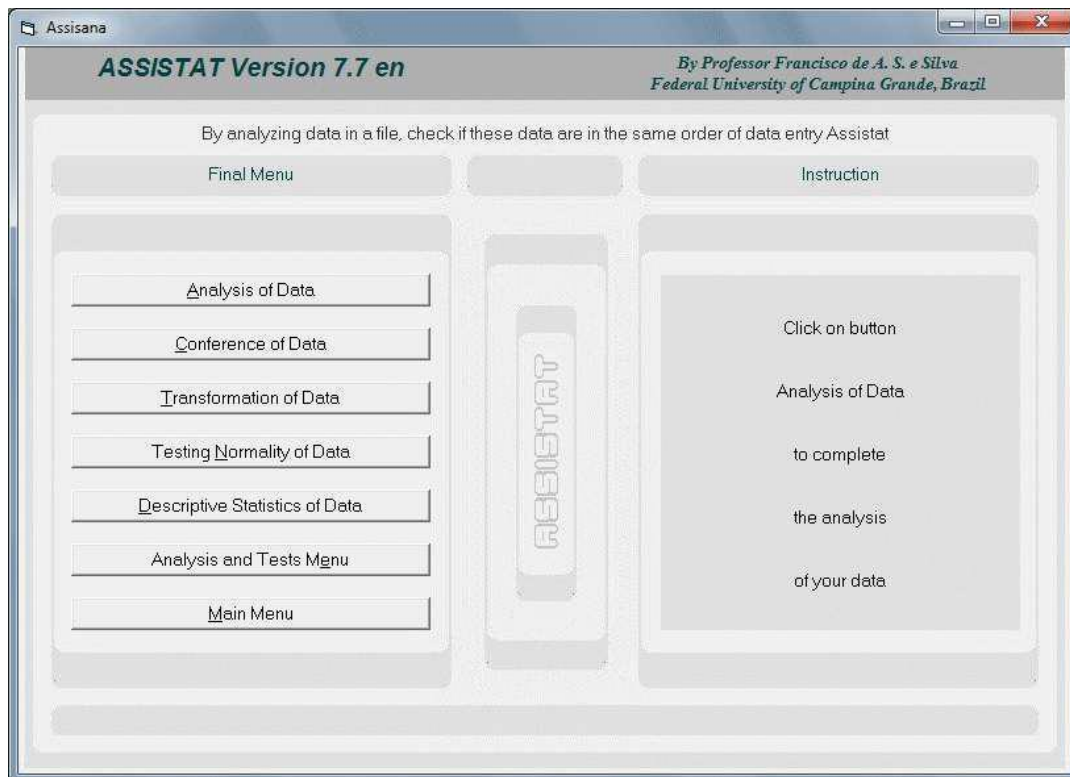


Figure 3. Final menu.

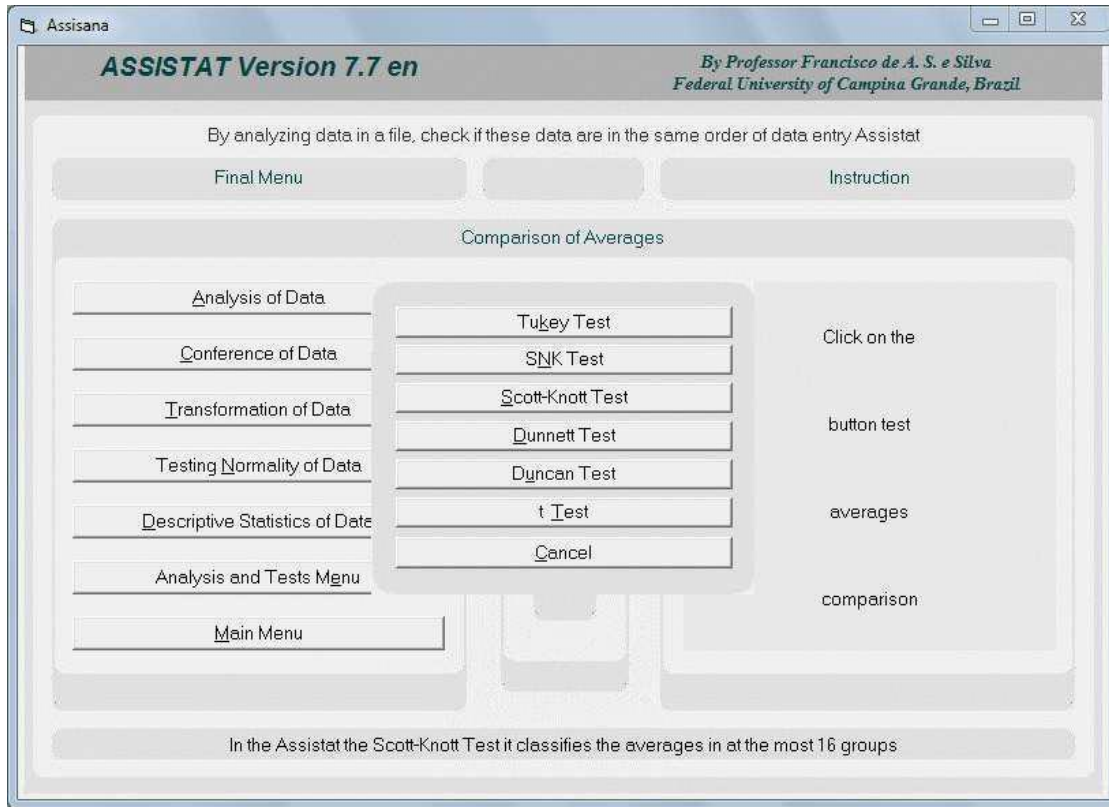


Figure 4. Tests of comparison of means.

Table 1. Content of copper (ppm) in sugarcane leaves.

Block	Treatment				
	A	B	C	D	E
1	11.5	7.7	9.8	10.7	12.0
2	12.7	9.0	8.0	10.8	10.9
3	12.6	9.1	7.4	10.2	10.3
4	12.2	8.6	9.5	9.6	9.8
5	10.4	8.8	8.3	9.8	9.4
6	12.0	8.6	8.9	10.1	9.5
7	12.2	8.4	10.5	11.1	9.7
8	8.5	8.4	10.4	11.5	10.1

Source: Campos (1984:66).

software allows the selection of other tests (Figure 3). The study used data of one randomized block experiment (Campos, 1984) and one factorial experiment (Snedecor and Cochran, 1979). In addition, regression was used in the analysis of variance for quantitative treatments (Gomes, 2009). The data of these experiments are respectively presented in Tables 1 to 3.

## RESULTS AND DISCUSSION

The data in Table 1, for the experiment in randomized

blocks, were analyzed and the results of the analysis of variance (ANOVA) are shown in Table 4. The effect of blocks was not significant and, even if it was, in field experiment the comparison of means of this effect is irrelevant and, therefore, dispensable. On the other hand, the effect of treatment was significant; thus, there is difference between the treatments, which indicates the necessity of comparison of means, presented in Table 5.

For the data of Table 2, factorial experiment, the result of the ANOVA is presented in Table 6. All the effects

**Table 2.** "Pajade guisante" yield (pounds/100 "morgen" parcelle) of three varieties.

Variety	Spacing	Block			
		1	2	3	4
I	4	56	57	56	53
I	8	45	50	60	62
I	12	43	50	53	68
II	4	46	65	53	67
II	8	60	61	48	60
II	12	50	60	55	73
III	4	45	63	60	77
III	8	48	60	61	77
III	12	66	58	50	65

Source: Snedecor and Cochran (1979;435).

**Table 3.** Maize production, in kg/plot, of an experiment of fertilization.

Block	Treatment (kg/hectare of P <sub>2</sub> O <sub>5</sub> )				
	0	25	50	75	100
1	3.38	7.15	10.07	9.55	9.14
2	5.77	9.78	9.73	8.95	10.17
3	4.90	9.99	7.92	10.24	9.75
4	4.54	10.1	9.48	8.66	9.5

Source: Gomes (2009), page 232.

**Table 4.** ANOVA for the data of Table 1.

Variation source	Degrees of freedom	Mean square
Blocks	7	0.6141 <sup>ns</sup>
Treatments	4	10.7919**
Error	28	1.0189

\*\* Significant at 0.01 probability level ( $p < 0.01$ ), \* Significant at 0.05 probability level ( $0.01 \leq p < 0.05$ ), ns Not significant ( $p \geq 0.05$ ).

**Table 5.** Comparison of means referring to the data of Table 1 by Tukey test at 0.05 probability level.

Treatment	Mean
1	11.5125 <sup>a</sup>
2	8.5750 <sup>c</sup>
3	9.1000 <sup>bc</sup>
4	10.4750 <sup>ab</sup>
5	10.2125 <sup>ab</sup>

Means followed by the same letter do not differ statistically.

were significant; however, the important ones are the effects of variety, spacing and the interaction between

them.

Table 7 shows the comparison of means for the effects

**Table 6.** Results of the ANOVA for the data of Table 3.

Variation source	Degrees of freedom	Mean square
Variety (V)	2	513.6944**
Spacing (S)	2	77.5278*
Interaction V x S	4	191.3611**
Treatments	8	243.4861**
Blocks	3	85.2130**
Error	24	17.6713

\*\* Significant at 0.01 probability level ( $p < 0.01$ ), \* Significant at 0.05 probability level ( $0.01 \leq p < 0.05$ ), ns Not significant ( $p \geq 0.05$ ).

**Table 7.** Comparison of means for the effects of variety and spacing by Tukey test at 0.05 probability level.

Variety	Mean	Spacing	Mean
I	51.3333 <sup>c</sup>	4	55.2500 <sup>b</sup>
II	57.6667 <sup>b</sup>	8	57.8333 <sup>abu</sup>
III	64.4167 <sup>a</sup>	12	60.3333 <sup>a</sup>

Means followed by the same letter in the column do not differ statistically.

**Table 8.** Comparison of means for the interaction variety x spacing by Tukey test at 0.05 probability level.

Variety	Spacing 4	Spacing 8	Spacing 12
I	47.5000 <sup>bd</sup>	50.7500 <sup>abd</sup>	55.7500 <sup>ba</sup>
II	62.2500 <sup>aa</sup>	58.5000 <sup>abd</sup>	52.2500 <sup>bd</sup>
III	56.0000 <sup>ab</sup>	64.2500 <sup>ad</sup>	73.0000 <sup>am</sup>

Means followed by the same letter in columns or rows do not differ statistically (Column = lowercase letters; Rows = uppercase letters).

of variety and spacing. The three varieties are different, while the spacing 8 does not differ from the spacings 4 and 12. The spacings 4 and 12, however, are different. Table 8 shows the result of the crossed Tukey test, also known as A inside B and B inside A. It is observed that the varieties II and III do not differ at the spacings 4 and 8, and that the varieties I and II do not differ at the spacing 12. On the other hand, the three spacings behave differently for each one of the three varieties.

Table 9 shows the result of the regression in the analysis of variance for the data of Table 3. In this case, the important effects are only the effects of regression, and it is observed that the effects of linear, quadratic and cubic regressions were significant. This means that these three types of equations express the dependence between the response variable (Maize production) and the quantitative treatments (0, 25, 50, 75 and 100 kg/ha

of P<sub>2</sub>O<sub>5</sub>). Nevertheless, in a case like this, the second-degree equation may be accepted, but the third-degree equation is actually the one that best expresses this

dependence, because it has higher coefficients of correlation ( $r$ ) and determination ( $R^2$ ).

The next step of the analysis was to obtain the equations for the significant effects. In the Assistat software, these equations are provided as part of the results. Table 10 shows the first-, second- and third-degree equations (linear, quadratic and cubic) and their respective coefficients of correlation ( $r$ ) and determination ( $R^2$ ).

Results similar to those of the present study were obtained, using the Assistat software, by Borcioni (2016), Brandelero et al. (2015), Campos et al. (2014), Carvalho et al. (2016), Dias et al. (2014), Diniz Neto et al. (2014), Gassen et al. (2014), Mendonça et al. (2014), Modesto et al. (2014), Oliveira and Albrecht (2014), Schwalbert et al. (2016), Silva et al. (2015), Silva et al. (2016), Sousa et al. (2014) and Zeist et al. (2014).

These cited papers and many others that were reviewed and that used the Assistat software confirm its acceptance, which points to its functionality and efficiency

**Table 9.** Regression in the analysis of variance for the data of Table 3.

Variation source	Degrees of freedom	Mean square
Linear regression	1	40.6426**
Quadratic regression	1	21.2791**
Cubic regression	1	9.2256**
Fourth-degree regression	1	1.0726 <sup>ns</sup>
Treatments	4	18.0550 <sup>ns</sup>
Blocks	3	0.9116
Error	12	0.9094

\*\* Significant at 0.01 probability level ( $p < 0.01$ ), \* Significant at 0.05 probability level ( $0.01 \leq p < 0.05$ ), ns Not significant ( $p \geq 0.05$ ).

**Table 10.** Regression equations in the analysis of variance and their respective coefficients of correlation ( $r$ ) and determination ( $R^2$ ).

Degree	$y = a + b.x + c.x^2 + d.x^3$				$r$	$R^2$
	a	b	c	d		
1	6.42250	0.04032	0	0	0.75017	0.56276
2	5.18964	0.13895	-9.86286E-04	0	0.92596	0.85740
3	4.70939	0.27662	-4.82826E-03	2.56133E-05	0.99255	0.98516

$y$  = Maize production (kg/plot);  $x$  = Quantitative treatments (0, 25, 50, 75 and 100 kg/ha of  $P_2O_5$ ).

in the analysis of experimental data of agricultural research, since most of the papers are from this area. Besides these references, many others in which the software was used can be found in various journals; only a small part of them were cited here.

## Conclusion

Considering the results of the present study and the utilization in published papers, it can be concluded that the Assistat software version 7.7 is functional and efficient in the analysis of experimental data of agricultural research. It is available for download at: <http://www.assistat.com>

## Conflict of interests

The authors have not declared any conflict of interests.

## REFERENCES

- Borcioni E, Mógor AF, Pinto F (2016). Aplicação de ácido fúlvico em mudas influenciando o crescimento radicular e produtividade de alface americana. *Rev. 510 Ciênc. Agron.* 47(3):509-515, jul-set.
- Brandelero EM, Araújo AG, Ralisch R (2015). Mobilização da cobertura vegetal na linha de semeadura e sua influência na temperatura e conteúdo de água no solo e na emergência do milho. *Eng. Agríc.* 35:1 Jaboticabal jan./fev.
- Campos H (1984). *Estatística experimental aplicada à experimentação com cana-de-açúcar*. São Paulo: FEALQ 292 p.
- Campos PF, Alves Júnior J, Casaroli D, Fontoura PR, Evangelista AWP (2014). Variedades de cana-de-açúcar submetidas à irrigação suplementar no cerrado goiano. *Eng. Agríc.* 34:6.
- Carvalho CA, Silva J, Alves CZ. (2016). Envelhecimento acelerado em sementes de mogno. *Rev. Ciênc. Agron.* 47(4):691-699, out-dez.
- Cox DR, Reid N (2000). *The Theory of the Design of Experiments*. Chapman and Hall/CRC, Boca Raton. 323 p.
- Dias VO, Alonço AS, Carpes DP, Veit AA, Souza LB (2014). Tamanho amostral para ensaios de mecanismos dosadores de sementes de milho em esteira carpetada. *Eng. Agríc.* 34:5.
- Diniz Neto MAD, Silva IF, Cavalcante LF, Diniz BLMT, Silva JCA, Silva EC (2014). Mudanças de oiticica irrigadas com águas salinas no solo com biofertilizante bovino e potássio. *Res. Bras. Eng. Agríc. Ambient.* 18(1):10-18.
- Gassen JRF, Alonço AS, Baumhardt UB, Mateus P, Bellé MP, Bonotto GJ (2014). Resistência específica à tração na operação de escarificação do solo em camadas de forma simultânea. *Res. Bras. Eng. Agríc. Ambient.* 18(1):116-124.
- Gomes FP (2009). *Curso de estatística Experimental*. 15ed. Piracicaba: FEALQ 451p.
- Mendonça VZ, Mello LMM, Pereira FCBL, Silva JOR, Yano EH (2014). Produção de milho para silagem consorciado com forrageiras na integração lavoura-pecuária. *Eng. Agríc.* 34:4.
- Modesto LR, Lustrri EA, Vasconcelos FJ, Soares FJVL, Silva RPF, Fluminhan Junior A (2014). Seleção recorrente fenotípica visando à obtenção de uma variedade sintética de milho (*Zea mays* L.). *Colloq. Agrar.* 10(2):20-31.
- Montgomery DC (2008). *Design and Analysis of Experiments*. Ed. Wiley; 7th edition.
- Oliveira DP, Albrecht JM (2014). Avaliação de substratos na germinação de *Plathymenia reticulata* Benth. *Rev. Biodiversidade* 10:1.
- Schwalbert R, Maldaner J, Schwalbert RA, Silva LOS, Steffen GPK, Steffen B (2016). Optimization of culture media for desmodium incanum micropropagation. *Afr. J. Agric. Res.* 11(12):1062-1070.
- Silva LMM, Cavalcanti Mata MERM, Duarte MEM (2015). Teor de água limite para criopreservação de sementes de romã (*Punica granatum* L.). *Eng. Agríc.* 35:2 Jaboticabal mar./abr.

- Silva RC, Grzybowski CRS, Panobianco M (2016). Vigor de sementes de milho: influência no desenvolvimento de plântulas em condições de estresse salino. *Rev. Ciênc. Agron.* 47(3):491-499..
- Silva FAS, Azevedo CAV (2006). A New Version of The Assistat-Statistical Assistance Software. In: *WORLD CONGRESS ON COMPUTERS IN AGRICULTURE*, 4, Orlando-FL-USA: Anais.. Orlando: Am. Soc. Agric. Biol. Eng. pp. 393-396.
- Snedecor GW, Cochran WG (1979). *Metodos estadísticos*. 6.ed. Tradução de J. A. Fuller. México: Companhia Editora Continental SA 704 p.
- Sousa CHC, Lacerda CF, Silva FLB, Neves ALR, Costa RNT, Gheyi HR (2014). Produtividade de sistemas de rotação algodão/ feijão-de-corda e girassol/ feijão-de-corda durante o processo de recuperação de um solo salino-sódico. *Eng. Agríc.* 34:5.
- Zeist AR, Zeist RA, Giacobbo CL, Resende JTV (2014). Produção de mudas de tomateiro utilizando substratos à base de substrato comercial (turfa fértil®), vermiculita expandida e fibra de coco. RAMVI, Getúlio Vargas 01:02.