Bibliometric research on the use of analysis of variance in seed germination studies

Gustavo D. Roesler[†], Josiane Rodrigues, Victor Augusto Forti

Universidade Federal de São Carlos (UFSCAR)

Resumo: *Estudos em fitotecnia com enfoque na investigação da germinação de sementes são desenvolvidos frente uma ampla gama de finalidades, sendo a análise estatística dos dados um componente imprescindível para confiabilidade e comprovação experimental. Entre os métodos estatísticos adotados, a análise de variância (ANOVA) seguida da aplicação de testes de médias caracteriza-se como a principal ferramenta utilizada na análise dos dados obtidos. Dentro deste contexto, o presente estudo foi realizado com o objetivo de avaliar quantitativamente o uso da ANOVA na produção científica envolvendo estudos com germinação de sementes, utilizando para isso uma análise bibliométrica para o período de 2000 a 2020, considerando a base de dados Web of Science. A estratégia de pesquisa consistiu na indexação de termos de busca para a ANOVA por operadores booleanos e de truncagem, restringindo a busca à artigos originais. Avaliou-se as relações das publicações entre áreas de conhecimento, evolução anual da produção científica, países e revistas nas quais os artigos foram publicados e instituições de pesquisa responsáveis pela realização dos estudos, observando predomínio de publicações desenvolvidas sob o eixo de agricultura (52,5%), crescimento da produção científica utilizando-se da ANOVA nas últimas duas décadas, e notório destaque da produção científica atribuída ao Brasil e revistas brasileiras, responsáveis por 46,7 e 44% das publicações do período, respectivamente.*

Palavras-chave: *Agricultura; análise estatística de dados; ANOVA; estudo de revisão.*

Abstract: *Studies in phytotechny focusing on the investigation of seed germination are developed for a wide range of purposes and the statistical analysis of data representing an essential component for reliability and experimental proof. Among the statistical methods adopted, the analysis of variance (ANOVA) followed by the application of means tests is characterized as the main tool used in the analysis of the data obtained. In this context, the present study aimed at evaluating quantitatively the use of ANOVA for scientific production involving seed germination studies, using for this purpose a bibliometric analysis for the period from 2000 to 2020, considering the database Web of Science. The search strategy consisted in indexing search terms for ANOVA using boolean and truncation operators, restricting the original articles. The relations of the publications among the areas of knowledge were evaluated, as well as the annual evolution of scientific production, countries, and journals in which the articles were published and research institutions responsible for conducting the studies. It was observed a prevalence of publications developed under the agriculture focus (52.5%), growth of the scientific production using ANOVA in the past two decades, and a notorious emphasis on scientific production attributed to Brazil and Brazilian journals, responsible for 46.7 and 44% of the publications in the period, respectively.*

Keywords: *Agriculture; statistical analysis of data; ANOVA; review study.*

Introduction

Studies on seed science are developed for a wide variety of goals that permeate plant science, for instance phytotechny, ecology, genetics and weed management, in which variables such as: time, uniformity, speed, synchrony and percentage of germination are usually considered (RANAL and SANTANA, 2006).

Regarding seed germination, there are several methodologies which can be used in data analysis, as the example of the analysis of variance (ANOVA), germination index, methods of classic regression (linear or non-linear regression), non-parametric tests, generalized linear models and survival analysis techniques (AZEVEDO, 2016).

[†] Autor correspondente: gu_roesler@outlook.com.

Among the experimental methods cited, ANOVA followed by the application of mean comparison tests is characterized as one of the most adopted statistical models (CARVALHO et al., 2018), and its continuous use implies the consolidation of the method to a large part of the phytotechnical tests involving seed analysis.

ANOVA is a statistical procedure which compares, within an experiment, the variation derived from the treatments with the variation of error, or residual. The method is based on a normal linear model that emphasizes the fundamentality of repetition, randomization, and local control in experimental efficiency (FISHER, 1934). Therefore, for the appropriate use of the technique, it is necessary to meet some assumptions, namely: homogeneous experimental error variances, independent and normally distributed errors, and the effect of additivity between treatments and blocks, this last one in the cases where the use of the blocking principle is necessary (SOKAL and ROHLF, 1995).

According to the Central Limit Theorem, replication on plates, considered as the experimental units in germination assays, provides a statistical evaluation of the standard errors of the germination means, which are, therefore, appropriate for ANOVA and for means tests, since the means of the replicated plates tend to stay close to a normal distribution of the errors (GIANINETTI, 2020).

Given the exposed, the aim of this study was evaluating quantitatively the use of ANOVA in the scientific production involving seed germination, employing for this a bibliometric analysis for the period from 2000 to 2020 considering the database Web of Science.

Methodology

Data collection was performed based on the main collection of the database of the platform Web of Science, which encompasses the main academic journals, books and annals referring to the world scientific production (offering access to six citation databases: *Science Citation Index Expanded* – SCI-EXPANDED, *Social Sciences Citation Index* – SSCI, *Arts & Humanities Citation Index* – A&HCI, *Conference Proceedings Citation Index-Science* – CPCI-S, *Conference Proceedings Citation Index-Social Science & Humanities* – CPCI-SSH and *Emerging Sources Citation Index* – ESCI). The study was performed for the period from 2000 to 2020.

All records containing the terms *"analysis* of* variance*" and "germination"* in the title, abstract and/or keywords were selected. The "quote" (") truncation operator limits the search to adjacent terms in the text, fixing the sequence, in this order, and without any other word for the expression of interest, and the asterisk (*) at the end of the word allows setting a prefix, enhancing the results found. Furthermore, the Boolean logical operator *"and"* was applied in the composition of the search terms, which allows the combination of two terms, in this case, *"analysis of variance"* and *"germination"*.

Following the methodology usually employed in bibliometric studies, the search was restricted to original articles published in periodicals, excluding the review articles and annals of scientific meetings, since the peer review criteria adopted for the publication of original scientific articles result in great reliability regarding the experimental method, reproducibility, and significance (DARCH and UNDERWOOD, 2005). Thus, 348 articles were selected. Each of the works was individually analyzed, aiming at sorting publications that indeed referred to the application of the statistical methods of interest in seed germination studies, reaching a final number of 289 articles. The search was performed in January 2021.

Once the search terms were defined, the bibliometric analysis was developed exclusively quantitatively. For this analysis, the following variables were considered: relations of publications among areas of knowledge, annual evolution of the scientific production, countries, and journals in which the articles have been published and research institutions responsible for performing the studies.

The limitations resulting from indexing single terms for data collection, as well as the use of a single search base, are aspects that limit the sample field of the research, being largely important the performance of works with greater scope to evaluate more comprehensively the analysis of statistical tools in the scientific production regarding seed science.

Results and Discussion

Analyzing the relationship among the research fields that encompass the publications which have used ANOVA to interpret the seed germination data, it was verified that, of the 289 selected articles, around 52.2% were characterized as studies developed under a central focus in agriculture, followed, in order of number of publications, by the topics: plant science (12.5%), forestry (8.3%), environmental sciences and ecology (5.9%), food science and technology (4.8%), science and technology (3.5%), biotechnology and applied microbiology (3.1%), biomedicine (2.8%), chemistry (2.1%) and biochemistry and molecular biology (2.1%). In this sense, the set of the 10 priority research areas encompasses 97.3% of the publications, the rest being identified in the areas of engineering, medicine and energy and fuels (data not presented).

Concerning the research areas which have adopted ANOVA for the analysis of data derived from seed germination studies, it was observed that the large use related to the area "agriculture" translates the statistical composition usually adopted in these studies, since in the agricultural sciences, as well as in seed science, the use of this tool and its derivations, such as consolidated statistical methods, is presented as an assiduous aspect and extensively applied in scientific experiments (NETTER et al., 1996; CARVALHO et al., 2018).

Regarding the evolution of scientific production over the years, it is verified that, from 2008, there was an increase in the number of publications that have used ANOVA in seed germination studies, reaching the highest index in the year 2020 (Figure 1), an aspect which corroborates the statement that this is a usual way of analyzing data in agronomic experiments (GIRARD et al., 2009). Therefore, it is also evidenced that this aspect is also a reflection of the characteristic professional training of researchers, which inherently encourages the adoption of ANOVA for understanding the biological phenomena studied, which is a profoundly valid perspective and supported by the structuring of the analysis, as long as it is well suited.

Figure 1: Absolute frequency of the evolution of the scientific production regarding the application of the analysis of variance (ANOVA) in seed germination studies, in the period from 2000 to 2020.
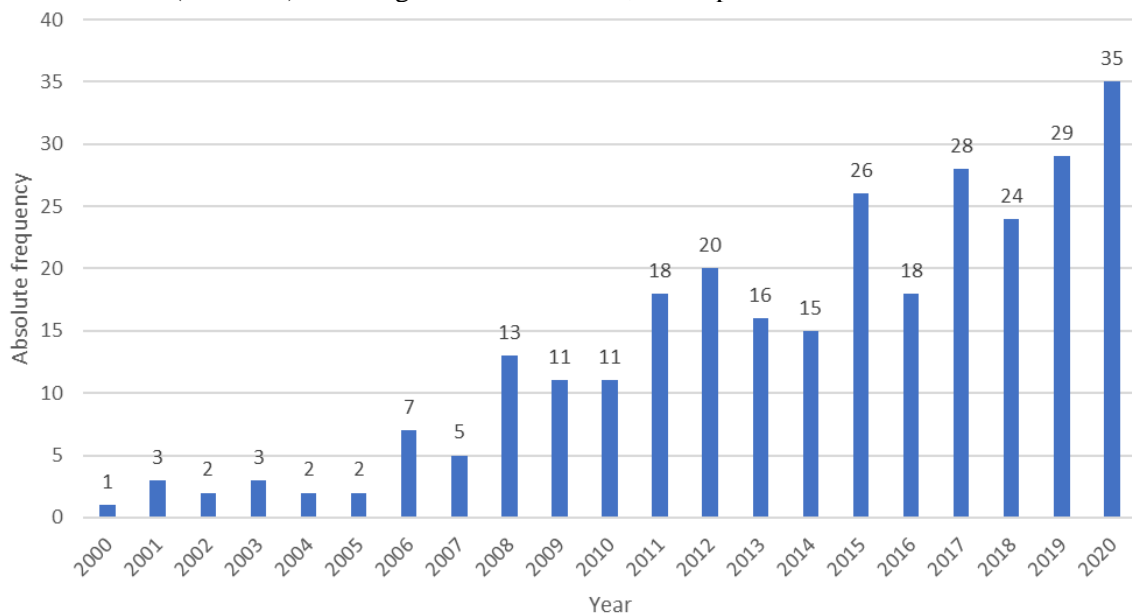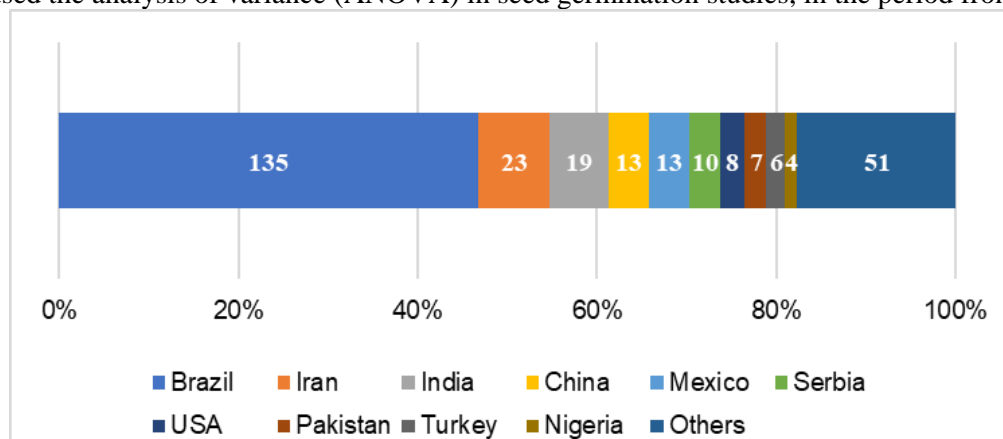

Fonte: Authors.

Addressing the distribution of publications regarding the countries responsible for producing the articles it was observed a total of 46 countries. Nevertheless, it is worth highlighting that 24 of them correspond to only one publication. Analyzing the 10 main countries responsible for the scientific production of germination studies involving the application of ANOVA, the great highlight attributed to

Brazil is notorious, encompassing 46.7% of the total of published articles, 6 times more scientific publications than Iran (7.9%), the second country in number of publications (Figure 2).

Figure 2: Representation of the 10 most productive countries in the publication of scientific articles that have used the analysis of variance (ANOVA) in seed germination studies, in the period from 2000 to 2020.

Fonte: Authors.

This fact demonstrates the great scientific collaboration of Brazil for studies related to seeds, so that, analyzing under a broader perspective, this finding is supported by the characteristic profile of the Brazilian scientific production in agricultural sciences, which is among the main countries with the greatest number of articles in this large research field (CAÑAS-GUERRERO et al., 2013). Furthermore, these same authors highlight the remarkable growth of the research works in agronomy in countries such as India and China, the third and fourth countries which have most produced between 2000 and 2020 considering the established topics, respectively.

Regarding the journals which have published the articles under analysis, the 289 articles have been published in 151 different journals, 104 of which correspond to the publication of a single article. Thus, the scientific journals that have published 3 or more articles for the period are presented in Figure 3. Considering only the journal with 3 or more publications, it was observed 135 articles, corresponding to 80 publications indexed in Brazilian journals, which occupy 17 positions among the 22 most active journals (Figure 3).

In this regard, it is important to highlight that the small number of articles linked to high visibility journals in the field of seed technology, as the example of the Journal of Seed Science, refer, besides the definition of the terms, to the fact that these journals have been recently indexed in the database Web of Science. Under this consideration, it is observed that the publications referring to the Journal of Seed Science are listed in the database Web of Science from 2014, since the journal was submitted to evaluation in the platform in 2013, and, indeed, it was indexed in the database in 2016 (ABRATES, 2019); therefore, for the period analyzed (2000 – 2020), only 7 years (2014 – 2020) of the journal's activity were analyzed.

Considering all 151 journals responsible for the publication of the 289 articles related to the ANOVA of the germination data, it was verified that 28 are Brazilian journals, even if they present an international character, and 123 refer to the other countries (Figure 4A). Nonetheless, even with a prevalence, in absolute numbers, of journals from other countries, considering only the articles published in the Brazilian journals, they comprise 128 publications, in other words, approximately 44% of the total analyzed, highlighting the strong participation of the scientific activity in Brazil in this research field (Figure 4B).

Furthermore, analyzing the bibliometric data under the perspective of organizations/institutions that have developed the published articles, a large Brazilian participation was verified again, since, of the 10 organizations that have most published in the period, 7 are Brazilian universities or research centers, which total 53 publications (data not presented).

Figure 3: The main journals which have published articles related to the analysis of variance (ANOVA) in seed germination studies, in the period from 2000 to 2020.
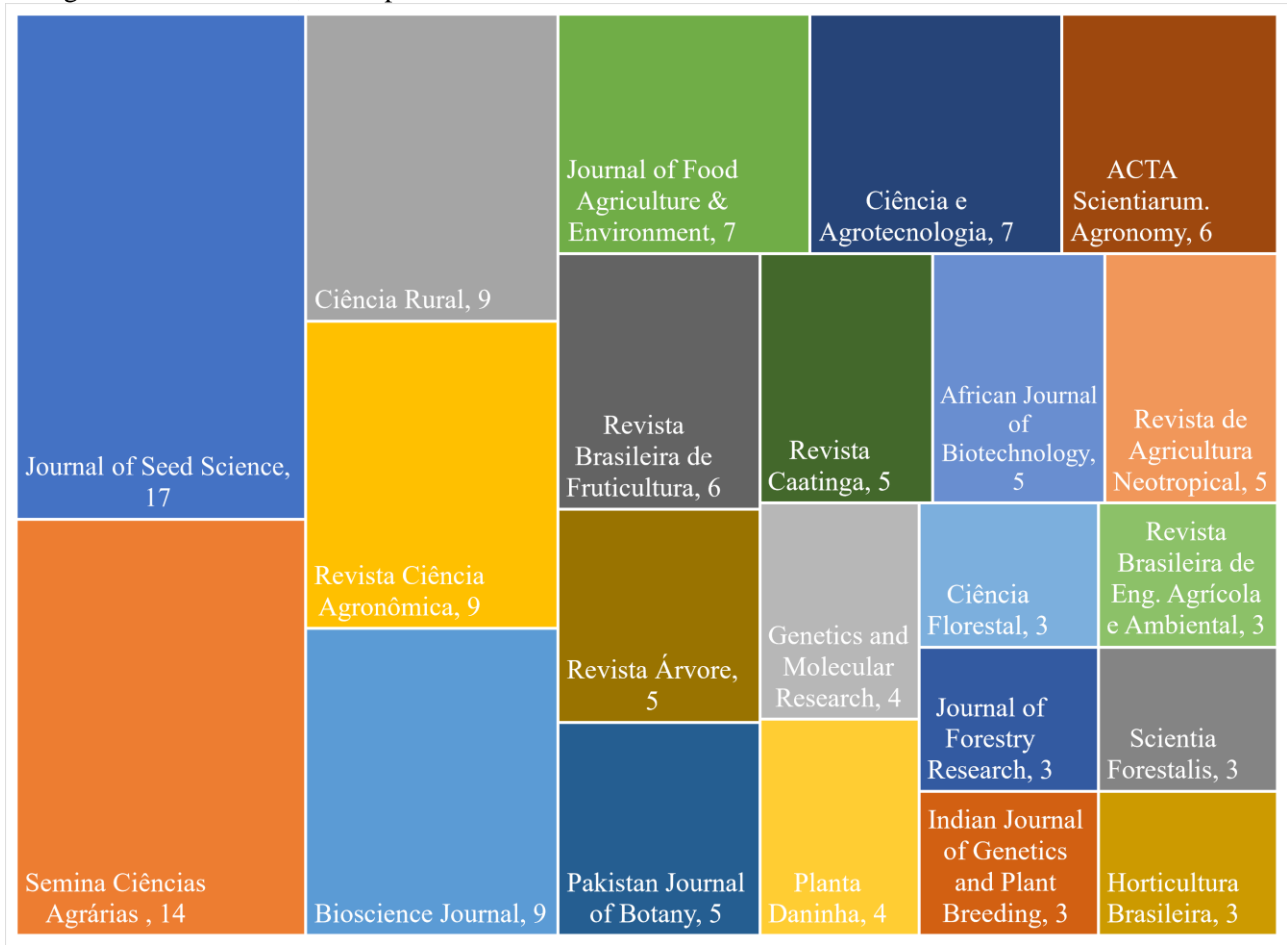
Fonte: Authors.

Figure 4: Proportion of journals responsible for the publication of the 289 articles using the analysis of variance (ANOVA) in seed germination data, in the period from 2000 to 2020 (A); Proportion of publications linked to Brazilian journals or journals from other countries considering the total of 289 articles using the analysis of variance (ANOVA) in seed germination data, in the period from 2000 to 2020 (B).
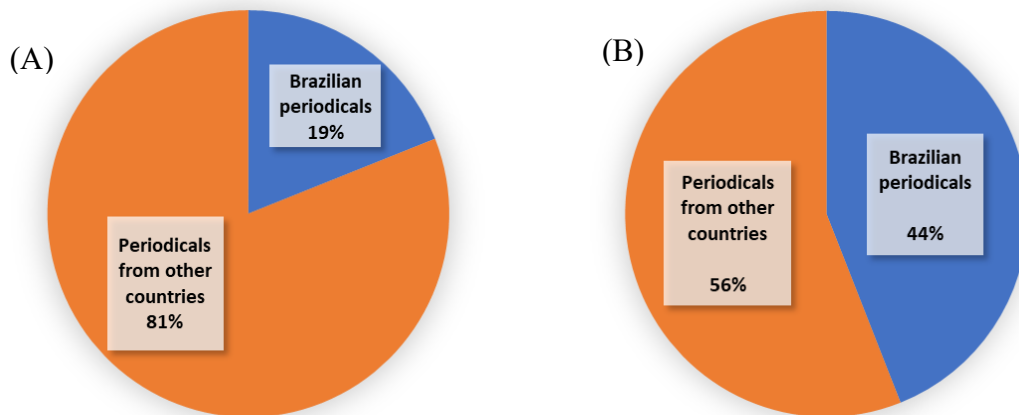
Fonte: Authors.

Therefore, considering the position of the journals by the number of articles, a major participation of Brazil is observed in studies with seeds, also highlighted by the presence of consolidated research institutions. In this sense, it is evidenced that this high number of produced articles is directly related to educational institutions and postgraduate programs, which historically have been an important source of scientific activity in the country (LYRA and GUIMARÃES, 2007), and which confirm the influence of the researchers' training profile in the adoption of ANOVA for seed science studies.

It is worth highlighting that, although ANOVA followed by the application of tests of means is a widely adopted technique in studies involving seed germination, several other methodologies can be used to analyze the behavior of a population of seeds regarding germination.

Conclusion

Regarding the use of ANOVA in the scientific production involving seed germination for the period from 2000 to 2020 and considering the database Web of Science, it was observed a prevalence of publications developed under the agriculture focus (52.5%), growth of the scientific production using ANOVA in the past two decades, and a notorious emphasis on scientific production attributed to Brazil and Brazilian journals, responsible for 46.7 and 44% of the publications in the period, respectively, an attribute that derives, besides the capabilities of the model itself, from the characteristic profile of the researchers involved in the area of seeds.

References

ABRATES – Associação Brasileira de Tecnologia de Sementes. Journal of Seed Science completa 40 anos em 2019. 2019. Disponível em: https://www.abrates.org.br/noticia/journal-of-seed-science-completa-40-anos-em-2019. Acesso em: 06 fev. 2021.

AZEVEDO, I. F. S. Modelos de dose-resposta com censura intervalar aplicados a dados de germinação de sementes. Tese (Doutorado em Ciências) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo. Piracicaba, 2016. 74p.

CAÑAS-GUERRERO, I.; MAZARRÓN, F. R.; POU-MERINA, A.; CALLEJA-PERUCHO, C.; DÍAZ-RUBIO, G. Bibliometric analysis of research activity in the "Agronomy" category from the Web of Science, 1997–2011. European Journal of Agronomy, v.50, p.19-28, 2013.

CARVALHO, F. J.; SANTANA, D. G. de; ARAÚJO, L. B. de. Why analyze germination experiments using Generalized Linear Models? Journal of Seed Science, v.40, n.3, p.281-287, 2018.

DARCH, C.; UNDERWOOD, P. G. Are pre-compiled citation indexes of peer-reviewed journals an adequate control for research quality? A case study of library and information science. South African journal of libraries and information science, v.71, n.1, p.1, 2005.

FISHER, R. A. Two new properties of mathematical likelihood. Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character, v.144, n.852, p.285-307, 1934.

GIANINETTI, A. Basic Features of the Analysis of Germination Data with Generalized Linear Mixed Models. Data, v.5, n.1, p.1-42, 2020.

GIRARD, L. H.; CARGNELUTTI FILHO, A.; STORCK, L. Erro tipo I e poder de cinco testes de comparação múltipla de médias. Revista Brasileira de Biometria, São Paulo, v.27, n.1, p.23-36, 2009.

LYRA, T. M. de P.; GUIMARÃES, J. A. Produção científica brasileira em comparação com o desempenho mundial em ciências agrárias. Planejamento de Políticas Públicas, v.1, n.30, p.141-162, 2007.

NETTER, J.; KUTNER, M. H.; NACHSTEIN, C. J.; WASSERMAN, W. Applied linear statistical models. Boston: Richard Irwin, 1996. 1408p.

RANAL, M. A.; SANTANA, D. G. de. How and why to measure the germination process? Revista Brasileira de Botânica, v.29, n.1, p.1-11, 2006.

SOKAL, R. R.; ROHLF, F. J. Biometry: the principles and practice of statistics in biological research. New York: W.H. Freeman, 1995. 776p.