The Use of Statistical Analyses in Papers and Graduate Programs in the Environmental Sciences Area in Brazil

Tatiel Venâncio Gonçalves 1
Thâmara Machado e Silva 2
Rafael Batista Ferreira 3
Werther Pereira Ramalho 4
Ronny José de Morais 5
Filipe Viegas de Arruda 6
Flávia Pereira Lima 7

ABSTRACT

Environmental issues emerge in complex dimensions, which require an interdisciplinary framework in Environmental Sciences. Due to the diversity in statistical methods, graduate programs need to upgrade in order to form environmental scientists. Here, we test the hypothesis that QUALIS A1 Journals in the Environmental Science area use more complex statistical analyses. We describe the tests offered by graduate programs with PhD degree in Environmental Sciences. A total of 33.5% of 1560 papers evaluated had no statistical analysis. A1 journals used more T-Test, Chi-Square and Mann-Whitney tests than B1 journals. We found no difference between the use of univariate, multivariate and Bayesian analyses. In Brazil, there are 37 graduate programs in Environmental Sciences, of which 10 do not offer statistics course. Among the 38 courses offered, 73.7% provide only univariate statistics and 34.2% provide multivariate statistics. We conclude that quality in papers is not dependable on the complexity of statistical analyses, but on their theoretical framework.

Keywords: Capes; QUALIS; Interdisciplinarity; Statistical Methods.

1 Doutorado em andamento em Recursos Naturais do Cerrado (Renac) pela Universidade Estadual de Goiás, UEG, Brasil. taticov17@hotmail.com
2 Doutorado em Ciências Ambientais pela Universidade Estadual de Goiás, UEG, Brasil. Professora no Centro Universitário Projeção, UniPROJEÇÃO, Brasil. thamaramachado.silva@gmail.com
3 Doutorado em andamento em Recursos Naturais do Cerrado (Renac) pela Universidade Estadual de Goiás, UEG, Brasil. Professor na Faculdade Metropolitana de Anápolis, FAMA, Brasil. rafael_g3bf@hotmail.com
4 Doutorado em andamento em Recursos Naturais do Cerrado (Renac) pela Universidade Estadual de Goiás, UEG, Brasil, com período sanduíche em Kansas State University. wertherpereira@hotmail.com
5 Doutorado em andamento em Recursos Naturais do Cerrado (Renac) pela Universidade Estadual de Goiás, UEG, Brasil. ronnybio@hotmail.com
6 Doutorado em andamento em Recursos Naturais do Cerrado (Renac) pela Universidade Estadual de Goiás, UEG, Brasil. filipeeco@gmail.com
7 Doutorado em andamento em Recursos Naturais do Cerrado (Renac) pela Universidade Estadual de Goiás, UEG, Brasil. Professora na Universidade Federal de Goiás, UFG, Brasil. linaflp@yahoo.com.br
Environmental issues, whose impacts are on biodiversity, on people’s welfare, on economy and politics, emerge from the complexity in natural and social dimensions, and from the interaction of both (Steele & Stier 2000; Liu et al. 2007). As consequence, researches in Environmental Sciences require an interdisciplinary approach to promote an effective conceptual integration of the natural and social sciences for the development of knowledge to be used in decision making processes (Rodela & Alasevic 2017).

The data to understand the complexity of the environmental issues are filled with uncertainties and present complex structured connections (Guttorp 2000). Considering this, statistics plays an important role in production, analysis, integration and dissemination of environmental data (El-Shaarawi & Teugels 2005), becoming a tool for understanding natural environments (Dowd et al. 2014; Marcionilio et al. 2016). It is also important in monitoring air pollution (Bellini et al. 2007), in preventing pollution in aquatic environments (Paroissin et al. 2016), in analyzing the effects of pollution on health (Lee et al. 2015), in evaluating and monitoring environmental impacts (McGeoch et al. 2015) and in environmental-related politics (Scott 2007).

The increasing use of statistics in Environmental Sciences comes from the need to perform analyses and objective assessments, substantiated in scientific knowledge (Clark & Gelfand 2006). Oftentimes, adopting new methods can reinforce the paradigm change and promote new scientific breakthroughs. The computational evolution caused an increased interest in intensive statistical methods, such as Generalized linear models, Non-linear and Bayesian models, in addition to resampling methods, such as permutation tests and bootstrap (Clark & Gelfand 2006). Therefore, research in Environmental Sciences addresses issues of high complexity that require an analytical framework for handling complex data. Although graduate programs do not yet provide a diversified formation in statistics, teaching advanced analyses would capacitate the future scientists for dealing with the environmental complexity (Piegersch & Edwards 2002; Butcher et al. 2007).

The scientific publishing process represents a competitive system of presenting ideas to a specific community. The number of published papers is way lower than the number of submitted
manuscripts, which creates competition. As consequence, the scientific community from higher quality journals expects originality and deepening questions. In Brazil, the quality of scientific journals is measured by the QUALIS, a measurement system of CAPES (Brazilian Federal Agency for the Improvement of Higher Education) to evaluate the scientific production of graduate programs. The scientific journals are classified in a system of quality grades (so called "strata") where A1 is the highest, followed by A2, B1, B2, B3, B4, B5 and C. This classification is divided in different areas evaluated by CAPES and uses many criteria, such as publication by Brazilian researchers, SJR (Scientific Journal Rankings) classification and presence in the SCOPUS database (CAPES 2013; Diniz-Filho et al. 2016). Therefore, the QUALIS classification have become a tool to measure the scientific journals quality, which allows to consult the grades of any journal of the CAPES Environmental Sciences area.

In this paper, we tested the hypothesis that higher impact journals in QUALIS system publish papers with more advanced statistical methods. We compared papers in A1 journals with B1 journals in the Environmental Sciences area. We expect that papers in A1 journals have more advanced statistics, such as multivariate and Bayesian analyses, than in B1 journals, on which descriptive statistics and univariate methods should be used more often. Moreover, we identified what statistical approaches and tests the graduate programs in Environmental Sciences have offered in their statistic courses.

**Material and Methods**

According to the QUALIS-2014 assessment (https://sucupira.capes.gov.br/sucupira/), there are 144 A1 journals and 311 B1 journals in the Environmental Sciences area. We searched in the Web of Science database (www.isiknowledge.com) for all papers published in these journals in 2015. In total, we found 106942 papers of A1 journals and 49028 papers of B1 journals.

In each paper, we identified the statistical analyses performed, usually described in the Material and Methods and Results sections. Papers without any statistical analysis were classified in three categories: theoretical (approach on a given topic without statistical inference), revision or frequentist (when results are presented in percentage or only in a descriptive way). We did not include in our tests statistical analyses with frequency lower than 1% in papers. Posteriorly, we classified the analyses in the following categories: Univariate or Multivariate, Parametric or Nonparametric, and Bayesian. Each paper could be classified in more than one category, depending on the type of analysis and the number of analyses used. We used secondary statistical analyses only in their main nomenclature (e.g., Simple Linear Regression or Multiple Regression were both classified as “Regression”).

---

To identify the approaches and statistical tests that have been taught in graduate programs of Environmental Sciences in Brazil, we searched 38 courses offered in 27 graduate programs with PhD. We downloaded, in the Plataforma Sucupira system (https://sucupira.capes.gov.br/sucupira/), the statistics courses syllabuses, even if the term “statistics” was not been used to denominate the course. After reading the syllabuses, we found the following approaches: probability, descriptive statistics, univariate statistics, multivariate statistics and Bayesian statistics. We also described the main tests presented in the courses syllabuses.

We randomly selected 1% of all papers as sample to test this paper hypothesis (A1 = 1070; B1 = 490). Subsequently, we computed the statistical power of tests using the pwr package (Champely 2015) in the R software (R Core Team 2016). If the power was significant, then our sample size was sufficient and, therefore, we controlled type II errors (Cohen 1988). For each analysis or category, we used Chi-Square test to compare the frequency of papers between A1 and B1 journals to identify if differences in frequencies were significant. Finally, we used a frequentist approach to identify the approaches and statistical tests that have been taught in graduate programs of Environmental Sciences in Brazil. All analyses were performed in R software.

**Results**

We found the sample size (1% of total articles) as statistically sufficient to obtain conclusive results in all Chi-Square tests (pwr > 95%). Among the 1560 papers analyzed, 523 did not present any statistical analysis. 32.9% (n = 353) of A1 papers and 34.7% (n = 170) of B1 papers had no statistical test. In papers with statistical analysis, 96.7% of A1 papers used any parametric test, while B1 journals had parametric tests in 98.4% of papers, indicating that B1 journals used more parametric tests than A1 ($\chi^2 = 5.41, gl = 1, p = 0.02$). Whereas for nonparametric tests, the number was higher in A1 (13.5%) in comparison to B1 papers (7.2%) ($\chi^2 = 8.694, gl = 1, p = 0.003$).

The use of univariate analyses was similar in A1 and B1 papers (91.7% e 90.6%, respectively; $\chi^2 = 0.371, gl = 1, p = 0.543$). The same occurred in multivariate analyses, that were similarly used in both journals (A1 = 11.3%, B1 = 12.5%; $\chi^2 = 0.311, gl = 1, p = 0.577$) and also in statistics using the Bayesian approach (A1 = 1.39%, B1 = 0.94%; $\chi^2 = 0.374, gl = 1, p = 0.541$).

The Student's T-Tests, ANOVA, Regression and Correlation were the most frequent, occurring in more than 20% of all papers analyzed (Figure 01). The T-Test was more frequent in A1 papers than in B1 papers (A1 = 32.79%, B1 = 22.74%; $\chi^2 = 10.013, gl = 1, p = 0.002$). The same
occurred for Chi-Square test ($A1 = 12.17\%$, $B1 = 7.02\%$; $\chi^2 = 7.787$, $gl = 1$, $p = 0.016$) and Mann-Whitney U test ($A1 = 6.53\%$, $B1 = 2.01\%$; $\chi^2 = 8.686$, $gl = 1$, $p = 0.003$) (Figure 01).

**Figure 01.** Frequency of occurrence of statistical analyses in A1 and B1 papers of the Environmental Sciences area. * $p < 0.05$.

![Frequency of occurrence of statistical analyses](image)

Source: Author (2016)

In papers without any statistical analysis, 39.4\% of A1 papers were theoretical, while 72.9\% of B1 papers fitted in this category ($\chi^2 = 51.707$, $gl = 1$, $p < 0.001$, Figure 02). Frequency in revision papers was similar in both strata ($A1 = 10.5\%$, $B1 = 9.4\%$; $\chi^2 = 0.144$, $gl = 1$, $p = 0.704$). Lastly, B1 papers used more frequentist statistics than A1 ($A1 = 49.3\%$, $B1 = 17.6\%$; $\chi^2 = 48.295$, $gl = 1$, $p < 0.001$).

**Figure 02.** Frequency of occurrence for each category of papers without statistics analyses in A1 and B1 papers of Environmental Sciences area. * $p < 0.05$.

![Frequency of occurrence for each category](image)

Source: Author (2016)

Among the 37 graduate programs with PhD evaluated in the Environmental Sciences area, 10 (27.02\%) have no statistic course. The majority of the 38 statistics courses presented univariate statistics.
(78.95%) and 47.37% proposed the study of multivariate methods, and also mostly addressed descriptive statistics approach (52.63%, Table 1). The most frequent statistical test in the syllabuses were: Regression (60.53%), ANOVA (39.47%), Correlation (34.21%) and Ordination methods (34.21%). Analyses such as T-Test (21.05%) and Chi-Square (7.89%) had low frequency in the syllabuses. The Bayesian approach was not present in any syllabus.

**Table 01.** Frequency of statistical tests present in syllabuses of graduate programs with PhD in the Environmental Sciences area (n = 38).

<table>
<thead>
<tr>
<th>STATISTIC</th>
<th>FREQUENCY (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Univariate</td>
<td>73.68</td>
</tr>
<tr>
<td>Multivariate</td>
<td>34.21</td>
</tr>
<tr>
<td>Descriptive</td>
<td>52.63</td>
</tr>
<tr>
<td>Probability</td>
<td>23.68</td>
</tr>
<tr>
<td>Bayesian</td>
<td>0.00</td>
</tr>
<tr>
<td>Regression</td>
<td>60.53</td>
</tr>
<tr>
<td>ANOVA</td>
<td>39.47</td>
</tr>
<tr>
<td>Correlation</td>
<td>34.21</td>
</tr>
<tr>
<td>T-Test</td>
<td>21.05</td>
</tr>
<tr>
<td>Chi-Square</td>
<td>7.89</td>
</tr>
<tr>
<td>Ordination</td>
<td>34.21</td>
</tr>
<tr>
<td>Cluster</td>
<td>28.95</td>
</tr>
<tr>
<td>Linear Models</td>
<td>7.89</td>
</tr>
<tr>
<td>Similarity</td>
<td>2.63</td>
</tr>
<tr>
<td>Non-parametric</td>
<td>28.95</td>
</tr>
</tbody>
</table>

Source: Author (2016)

**DISCUSSION**

We found a higher frequency of more traditional statistical methods in QUALIS A1 journals, in contrast to our predicted hypothesis. The quality of a scientific paper may not be necessarily related to the complexity of those analytical tools, but by how the scientific questions are developed and their hypotheses substantiated in a solid theoretical framework. The researcher, when applying a statistical test, need to master the statistics framework, such as factors, response variables, treatments and sampling design that will constitute his research (Bertoldo et al. 2007). Therefore, even facing the complexity of environmental questions and data, with well-delineated questions, the analytical tools required for answering them should be very simplistic.

Despite the journals categories do not differentiate regarding the use of statistics for both univariate and multivariate, we found dominance of univariate tests comparatively traditional (T-Test, Chi-Square and Mann-Whitney) in A1 journals. The increased use of interpretative statistical techniques was previously verified for the medical area between the 1950s and 1980s (Hayden 1983), reaching a proportional increase of 41% in 2005 (Hellem et al. 2007). In the financial area, published papers in
superior strata journals of QUALIS adopted more sophisticated statistical techniques (Cordeiro et al. 2014). Thereby, we understand that improvement in the use of inferential statistics nowadays is substantial, including in Environmental Sciences. Our results contradict Touchon and McCoy (2016), that found an increase in the use of Bayesian statistics and more sophisticated test in Ecology area. A possible explanation for this contradiction might be due to the higher area range in Environmental Sciences, which covers many other areas of knowledge.

The Brazilian graduate programs in Environmental Sciences provide a significant formation in univariate and descriptive statistics, despite the low frequency of T-Test and Chi-Square in the courses syllabuses. In general, such formation meets the demands of journals in the Environmental Sciences area. We consider that mastering more sophisticated statistical techniques can contribute to more robust interpretations of environmental issues, so it is important not to disregard the multivariate aspect of environmental data, due to their low temporal and spatial symmetry that can be used to model real situations (Guttorp 2003). Moreover, we consider high the percentage of graduate programs without statistics courses, which may indicate a mismatch between the top-level journals and the lack of disciplines that focus on these tests.

ACKNOWLEDGMENTS

We acknowledge CAPES, UEG and FAPEG for providing scholarships that enabled the elaboration of this study. We thank the Prof. Dr. João C. Nabout for the encouragement and all the suggestions offered. We also thank the RENAC professors, in special to each graduate student tutor, for the incentive in completing this project.

REFERENCES


The Use of Statistical Analyses in Papers and Graduate Programs in the Environmental Sciences Area in Brazil

Tatiel V. Gonçalves; Thâmara M. Silva; Rafael B. Ferreira; Werther P. Ramalho; Ronny J. Morais; Filipe V. Arruda; Flávia P. Lima


Os problemas ambientais emergem de dimensões complexas, exigindo um arcabouço interdisciplinar em Ciências Ambientais. Diante da diversidade de métodos estatísticos, os programas de pós-graduação precisam se atualizar para formar os cientistas ambientais. Nós testamos a hipótese que revistas de Qualis A1 na área de Ciências Ambientais usam estatísticas mais avançadas. Identificamos os testes estatísticos ofertados pelas disciplinas em programas de pós-graduação com doutorado em Ciências Ambientais. Dos 1560 artigos avaliados, 33.5% não apresentaram análise estatística. Revistas A1 utilizam mais Teste-T, Qui-Quadrado e Mann-Whitney que as B1. Não houve diferença no uso de análises univariadas, multivariadas e Bayesianas. No Brasil há 37 programas de pós-graduação em Ciências Ambientais, sendo que 10 não possuem disciplina de estatística. Das 38 disciplinas ofertadas, 73.7% oferecem estatística univariada e apenas 34.2% as multivariadas. Diante destes resultados, inferimos que a qualidade dos artigos não depende da complexidade das análises utilizadas, mas de seu arcabouço teórico.

Palavras-Chave: Capes; Qualis; Interdisciplinaridade; Métodos Estatísticos.

Submission: 17/11/2017
Acceptance: 20/08/2018