

Survey of statistical methods applied in articles published in Acta Scientiarum. Agronomy from 1998 to 2016

Omar Possatto Junior^{1*}, Filipe Augusto Bengosi Bertagna¹, Edicarlos Peterlini¹, Alessandra Guedes Baleroni¹, Robson Marcelo Rossi² and Hugo Zeni Neto³

¹Programa de Pós-graduação em Genética e Melhoramento, Universidade Estadual de Maringá, Av. Colombo, 5790, 87020-900, Maringá, Paraná, Brazil. ²Departamento de Estatística, Universidade Estadual de Maringá, Maringá, Paraná, Brazil. ³Departamento de Agronomia, Universidade Estadual de Maringá, Maringá, Paraná, Brazil. *Author for correspondence. E-mail: omar.pj@hotmail.com

ABSTRACT. Statistics is the main science by which researchers validate the results of scientific work, and the choice of an inadequate statistical method may lead to conclusions that are considered questionable by reviewers. This study had the objective of describing the characteristics of the statistical methods used in the papers published in Acta Scientiarum. Agronomy from 1998 to 2016 as part of a critical analysis of the journal to pinpoint possible failures in the application of these methods. All scientific articles ($n = 1,237$) published in the journal were surveyed, of which 54.1% addressed areas of crop production. The mean comparison methods were the most commonly used (75.5%) and, consequently, they represented the highest proportion of errors (60.8%) of the authors in the journal.

Keywords: statistical analysis; comparison of means; parametric methods.

Received on April 29, 2018.
Accepted on July 2, 2018.

Introduction

Population observation in the form of a census is costly and often impracticable. In view of this limitation, samples must be used for inferences on the population and, to draw these conclusions, researchers apply statistical techniques. Decision making or drawing conclusions is closely linked to statistics in most studies (Banzatto & Kronka, 2013; Barbin, 1993).

Statistics is the main science used by researchers to validate the results of scientific papers, regardless of the area addressed in the study. Due to this importance, the method should be carefully chosen according to the study data, in view of the risk of drawing conclusions that are considered questionable by the researcher's peers if an inappropriate statistical method is used (Montanhini Neto & Ostrensky, 2013; Conceição, 2008).

The choice of the statistical method should be based on a complete description of the applied technique, an appropriate use according to the scientific and statistical hypotheses and the tested treatments, as well as on the interpretations and conclusions that should be tuned to the results proposed by the method (White, 1979).

The analysis of variance (ANOVA) is one of the preferred methods in agricultural and biological experiments, which is used after determining the response variables in properly conducted experiments. Based on ANOVA results, it is possible to infer whether the evaluated treatments and blocks are significantly different (Banzatto & Kronka, 2013).

Prior to the analysis, some aspects must be taken into account, such as the basic assumptions (the distribution and independence of errors, the homogeneity of variances and the additivity of the model) and the nature of the treatment effects being fixed or random. The main objective of random-effect treatments is the estimation of the components of variance, which are highly relevant in plant and animal breeding. Conversely, the main goal of fixed-effect treatments is to compare these components by mean tests when they are qualitative or by regression analysis when they are quantitative (Bertoldo, Coimbra, Guidolin, Mantovani, & Vale, 2008; Bezerra Neto, Nunes, & Negreiros, 2002). These procedures should be applied with caution, since they may be suitable for certain types of treatments and inadequate for others.

Several statistical methods are available, although it is not uncommon to find articles in which an inappropriate use of the methods causes the researchers to draw misleading conclusions (Bezerra Neto et

al., 2002). According to Glickman et al. (2010), errors can also be a consequence of a lack of planning, mainly because some researchers only think about the statistical analysis after obtaining the experimental data.

In a census of 307 articles published in the Archives of Veterinary Science between 2000 and 2010, Montanhini Neto and Ostrensky (2013) found that the conclusions of only 32% of the studies were based on methods consistent with the treatment structure. Among articles of a journal classified as *Qualis A* (high quality by the official Brazilian system of the classification of scientific production) in the area of Agrarian Sciences that were published between 2000 and 2006, the mean comparison methods were applied correctly in only 22% of the 292 articles (Bertoldo et al., 2008).

In a review of articles published in Revista da Sociedade Brasileira de Zootecnia (SBZ) between 1984 and 1989, Cardellino and Siewerdt (1992) reported that only 24.6% of the mean comparison methods were applied correctly. They showed that the main errors consisted of an inappropriate use of ANOVA and mean comparison methods, the two most commonly used measures.

This study had the objective of describing the characteristics of the statistical methods applied in articles published in the journal *Acta Scientiarum. Agronomy* from 1998 to 2016 in order to quantify the possible failures in the application of these methods.

Material and methods

All scientific papers published in *Acta Scientiarum. Agronomy* from volume 20 in 1998 to volume 38, number 4 in 2016 were reviewed, resulting in a total of 1,237 articles.

The following information from each article was recorded: knowledge area, data source, number of years/growing seasons of the experiments, type of experimental design, treatment structure, number of replications, use of mathematical and/or statistical methods, use of assumptions, data transformation and their requirements, use of ANOVA, choice of statistical method (e.g., comparison of means, regression or multivariate analysis, nonparametric methods and descriptive statistics), methods of parameter estimation (such as least squares, likelihood and Bayesian methods), correlation analysis, use of a statistical program, and adequacy of the applied statistical methods.

The methodologies applied by the authors were classified as adequate, suboptimal or inadequate, based on the statistical concepts presented in the literature (Ferreira, 2008; Callegari-Jacques, 2003; Storck, 2000).

The statistics of an article were considered "adequate" when the applied statistical method consisted of the most appropriate procedure for the treatments described by the author. They were classified as "suboptimal" when the statistical method consisted of a convenient but not the most appropriate procedure and when the author used two statistical methods for the same dataset without the objective of comparing them.

The statistics were considered "inadequate" when a nonrecommended statistical method was applied to the study data set. For example, this included the application of a method of multiple mean comparisons in treatments of a quantitative nature in which the agricultural or biological interest is clearly perceived at intermediate levels of the response variable or in factorial experiments in which the marginal means of the factors were discussed without taking the possible interactions of the main effects into account.

All generated data were analyzed using descriptive statistics.

Results and discussion

The results of the area of knowledge, data source and years of experimental performance of the 1,237 articles are listed in Table 1. The majority of the articles published in *Acta Scientiarum. Agronomy* addressed areas of crop production (54.1%), followed by soil science (11.8%), crop protection (11.4%), and genetics and breeding (10.5%). These data show that the main research focus of the journal is crop production.

Most of the articles published in the journal used data from field experiments or controlled environments (54.2 and 41.9%, respectively). Articles based on literature reviews, questionnaires, sampling, and simulations represented 6.14% of the total number of published articles (Table 1). These results corroborate the results from a journal of veterinary science obtained by Montanhini Neto and Ostrensky (2013) and *Ciência Rural* by Lúcio et al. (2003), who observed that the majority of the results of articles were obtained in experiments in the field and controlled environments. The sum of percentages reached a value

of more than 100%, but it is worth mentioning that some studies included experiments in the field as well as in a protected environment, and the data were computed in both classes.

Table 1. Numbers and percentages of articles classified by areas of knowledge, data sources and years/growing season published in *Acta Scientiarum. Agronomy* between 1998 and 2016.

Descriptor	Number of papers	%
Field of knowledge		
Crop Production	670	54.10
Soil Science	146	11.80
Crop Protection	141	11.40
Genetics and Breeding	130	10.50
Agricultural Engineering	100	8.08
Microbiology	25	2.02
Biometrics, Modeling, and Statistics	9	0.73
Entomology	8	0.65
Physiology	8	0.65
Data Source		
Controlled environment	519	41.90
Field data	671	54.20
Sampling	33	2.67
Literature (review)	20	1.62
Simulation	14	1.13
Questionnaire	9	0.73
Experimental duration (years/growing seasons)		
1	977	87.50
2	83	7.44
3	20	1.79
4 or more	36	3.23

In the articles that presented experiments in the field or controlled environments that were conducted for only one year or one growing season (87.5% of the cases) (Table 1), the results might be altered if additional environments or years of cultivation were taken into consideration. On this account, the journal's publication norms were changed in 2010, requiring that the experiments be conducted in more than one environment to ensure more reliable data. For this and other reasons, the journal's paper quality increased and was classified as *Qualis A2* (high quality) in the quadrennial 2013-2016.

The results of the classification of papers with regard to design, treatment structure and number of replications are shown in Table 2. The complete block design with randomized treatments (RCBD) was used in 43.4% of the published articles, followed by the completely randomized design (CRD) in 31.4% of the articles, and Federer's lattice and block designs together accounted for 0.8% of the publications.

Table 2. Numbers and percentages of articles published in *Acta Scientiarum. Agronomy* between 1998 and 2016, classified by design, treatment structure and number of replications.

Descriptor	Number of papers	%
Design		
No Design	301	24.33
RCBD	537	43.41
CRD	389	31.45
Lattice	9	0.73
Augmented Federer blocks	1	0.08
Treatment structure		
Nested	651	58.33
Factorial	333	29.84
Split-plot design	115	10.30
Split-split plot design	6	0.54
Strip-plot design	9	0.81
Hierarchical design	2	0.18
Number of replications		
0 to 3	221	24.15
4 to 6	577	63.06
7 to 9	40	4.37
10 or more	77	8.42

Table 2 shows that no design was used in 24.3% of the articles. However, this category includes articles in which the authors did not state the design that was used and the articles about reviews, sampling, questionnaires and simulations, which required a non-experimental design.

In the area of animal science, Montanhini Neto and Ostrensky (2013) stated that the most popular design is the CRD, which is different from our findings. In research related to medicine, Conceição (2008) observed that a number of studies are based on patient samples consisting of groups of people. Therefore, the most appropriate designs for each study area are different since the experiments have different plots and sources of factors. Therefore, adequate designs are defined according to the specific requirements of the hypotheses to be tested.

Of all the articles, 1,116 (90.2%) described the structure of the treatments. Nested experiments represented 58.3% of the published articles (Table 2). Experiments in which the authors tested the combination of two or more factors, such as the crossed factorial with only one residue, the crossed factorial in split-plot design with two residues, the split-split plot design, or the strip-plot with three residues, represented 29.8%, 10.3%, 0.5%, and 0.8% of the publications, respectively.

Different results were reported by Bertoldo et al. (2008) in an evaluation of *Qualis A* journals, in which factorial experiments represented 65.3% and nested experiments represented only 34.7% of the published articles in the journals under study. The number of replications was stated in 74% of the articles published in the journal. Among these, between 4 and 6 replications were performed in 63.1% of the experiments, followed by experiments with up to 3 replications (24.1%), and then experiments with 7 to 9 or more than 10 replications, representing 4.3 and 8.4% of the articles, respectively (Table 2). In the animal and plant production area, Lúcio et al. (2003) observed that the mean number of replications was around four, as was similarly found in this research.

Replication is one principle of experimentation, since it allows estimation of experimental error. Therefore, an appropriate number of replications protects the precision of the experiment and treatment estimates. The higher the number of replications is, the better the experimental quality. However, in experiments with a high number of treatments, an increase in replications is not feasible since the size of the experimental area, biological material, seed quantity, and financial resources are limiting factors in determining the number of replications. With regard to the use of mathematical and statistical methods, the authors used statistical techniques to support their conclusions in 90.2% of the articles and in only 9.8% of the articles did the authors draw conclusions without statistical support, 16.5% of which were literature reviews (Table 3). Montanhini Neto and Ostrensky (2013) reported values of 66.5% and 33.6%, respectively.

Table 3. Numbers and percentages of articles, published in *Acta Scientiarum. Agronomy* between 1998 and 2016, classified according to the statistical methods, the test of basic assumptions, the data transformation and the level of significance required.

Descriptor	Number of papers	%
Applied methods		
Mathematical and statistical	1,116	90.22
Mathematical only	51	4.12
None	70	5.66
Level of significance for ANOVA		
1%	67	8.40
5%	728	91.23
10%	3	0.38
Test of assumptions		
Error normality	63	5.65
Homogeneity of variances	67	6.00
Additivity	1	0.08
Not mentioned	985	88.27
Data transformation		
Square root	82	59.42
Arcus Sinus	37	26.81
Log or ln	18	13.04
Box-Cox	1	0.72
Transformation required	41	29.71
Efficient transformation	24	17.39

Experimental data with statistical support indicate the reliability of the results and conclusions, but if this support is lacking, the analysis is deemed poor, incomplete or affected by the lack of knowledge on the

part of the scholar (Lúcio et al., 2003). It is possible to verify if a journal prioritizes publications based on experiments with statistical evaluations.

The analysis of variance was used in 64.5% of the articles (Table 3). According to Barbin (1993), it is the most frequently applied statistical method in experiments. For the journal *Archives of Veterinary Science*, (Montanhini Neto and Ostrensky, 2013) reported that the analysis of variance was used in 45% of the published articles. These results demonstrate that the procedure is still the most widely used method in experiments in agricultural and animal science.

The analysis of variance is a technique that consists of the partitioning of the total variance and degrees of freedom into parts attributed to known causes, which are the controlled factors (treatments) and to parts with unknown causes (residue) (Banzatto & Kronka, 2013; Fisher, 1971). The contribution of each part of the variance is highly relevant for researchers for helping to infer which treatments influence the resulting significance. However, to inspire confidence in the conclusions of the obtained results, it must be checked whether the data can be subjected to ANOVA by meeting the basic assumptions.

There are several methods available to verify the basic assumptions, such as the Shapiro-Wilk test and Lilliefors test, which examine the normality of error distribution (Campos, 1983; Guo, Alemayehu, & Shao, 2010). The Bartlett test tests the homogeneity of variances between treatments (Steel & Torrie, 1960), the sequence test verifies the error randomness (Beaver, Mendenhall, & Reinnhmuth, 1974), and the Tukey test of nonadditivity examines whether the effects of the mathematical model are additive and uses a minimum of 12 degrees of freedom (Snedecor & Cochran, 1967) for the residual analysis of variance.

The results regarding the use of basic assumptions and data transformation are presented in Table 3. The vast majority (88.2%) of the published articles did not mention the testing of at least one basic assumption, thus indicating that the results may be unreliable.

The main reasons why the authors fail to check the basic assumptions include a lack of knowledge about the tests and their importance, the confidence that the F (Snedecor) test is robust enough and that the data must not necessarily meet any basic assumptions, and not knowing what constitutes the assumptions. Similar results to these were found by Montanhini Neto and Ostrensky (2013), who reported that 81.2% of the published articles did not meet the basic assumptions.

When the data do fulfill the basic assumptions, some strategies can be used to proceed with the statistical analysis. The researcher can analyze the data through nonparametric methods or use data transformation methodologies (Lopes & Storck, 1995).

Methods of data transformation were observed in 11% of the articles. Of these, the method of extracting the square root was preferred by the authors in approximately 60% of the transformations. In 30% of the articles using transformations, their application was justified, and after the transformation, only 17% of the articles reported whether the procedure was efficient to meet the basic assumptions (Table 3).

Similar results were observed by Montanhini Neto and Ostrensky (2013) since 9.4% of the articles published in the reviewed *Journal of Animal Science* transformed the original data, while statistically this was necessary in only 3.3% of the cases. This shows that most authors apply data transformation without knowing if it is actually necessary but do so based on previous studies in which the authors applied transformations. It is worth mentioning that some authors forget to "bitransform" the data (return to the main unit) before presenting them in the results.

With regard to the levels of significance (α) presented in the articles, 91.2% of the studies used $\alpha = 5\%$, and 8.4% used $\alpha = 1\%$. The α used in the articles published in *Acta Scientiarum. Agronomy* are consistent with that of other articles published in agricultural journals. However, for Benjamin et al. (2017), an error probability of $\alpha = 5\%$ is high and decreases the credibility of new findings based on statistically significant results since the reproducibility of scientific studies is low. Based on this premise, they propose changing the standard level from $\alpha = 5\%$ to $\alpha = 0.5\%$.

The authors state two main benefits. First, the value of $\alpha = 0.5\%$ corresponds to a Bayes factor of approximately 14 to 26 in favor of H_1 (statistical significance), whereas $\alpha = 5\%$ would correspond to between 2.4 and 3.4. This method is used for the selection of models through the comparison of the *a posteriori* probabilities, and the model with the highest Bayes factor is preferable (Kass & Raftery, 1995).

The second benefit is that $\alpha = 0.5\%$ as a standard would reduce the type I error (false positive) rate to levels considered reasonable. In articles published with $\alpha = 5\%$, a type I error rate of more than 33% would decrease to 5% when using $\alpha = 0.5\%$. However, when setting $\alpha = 0.5\%$, the type II error (false negative) would become unduly high. Therefore, to overcome this inconvenience, one must increase the number of

replications by approximately 70% to ensure the statistical power. Parametric methods were used in most of the papers published in *Acta Scientiarum. Agronomy*, where multiple mean comparison methods represented 47.9% of all articles, followed by regression (31.1%) and multivariate techniques (17%) (Table 4).

Table 4. Classification of articles published in *Acta Scientiarum. Agronomy* between 1998 and 2016 with respect to the use of statistical methods of the comparison of means, regressions, multivariate analysis, nonparametric methods, and descriptive analysis.

Method	Number of papers	%		Number of papers	%
Mean comparison	592	47.86			
Tukey	447	75.51	<i>F</i> test	16	2.70
Fisher's <i>t</i> test	45	7.60	Bonferroni's <i>t</i> test	8	1.35
Duncan	25	4.22	SNK	8	1.35
Dunnett	22	3.72	Non-Orthogonal contrast	2	0.34
Orthogonal Contrast	18	3.04	Scheffé	1	0.17
Regression analysis	385	31.12			
Polynomial	322	83.64	Logistic	2	0.52
Non linear	36	9.35	Bisegmented	1	0.26
Multiple	24	6.23			
Multivariate analysis	210	16.98			
Scott-Knott	115	54.76	Multicollinearity	2	0.95
Clusters	37	17.62	ANACOVA	1	0.48
Repeated measures	12	5.71	Centroid	1	0.48
Principal components	12	5.71	Factor analysis	1	0.48
MANOVA	9	4.29	Discriminant analysis	1	0.48
Canonical variables	7	3.33	Canonical Discriminant analysis	1	0.48
Selection index	6	2.86			
Path analysis	4	1.90	Hotelling	1	0.48
Nonparametric analysis	36	2.91			
Chi-square	17	47.22	Friedman	1	2.78
Kruskal Wallis	13	36.11	Kolmogorov-Smirnov	1	2.78
Dunn	3	8.33	Mann-Kendall	1	2.78
Descriptive statistics	91	7.36			

Nonparametric methods represented 2.9% of the published articles, and the use of descriptive statistics was reported in 7.4% of the publications (Table 4). Montanhini Neto and Ostrensky (2013) observed that 44.1% of the articles used methods of mean comparison, 13.7% used regression analysis and 16.7% nonparametric methods, which is a five-fold frequency of nonparametric methods found in their journal of study compared to *Acta Scientiarum. Agronomy*.

In two journals on fruit crops, Cantuarias-Avilés and Dias (2008) evaluated the use of statistical methods and identified that parametric procedures were most commonly applied, as similarly found in this work. Moreover, descriptive and graphical statistical analysis represented 24% and 36% of the articles of the journals, respectively, which was different from those in *Acta Scientiarum. Agronomy* in which descriptive statistical analysis was used in 7.3%.

Among the mean comparison tests, Tukey's test was applied in more than 75% of the articles, followed by Fisher's *t* (7.6%), Duncan (4.2%), Dunnett (3.7%), orthogonal contrast (3%), *F* test (2.7%), Bonferroni *t* (1.3%), Student-Newman-Keuls (1.4%) and the Scheffé's test (0.2%). These results confirm Montanhini Neto and Ostrensky (2013), Cantuarias-Avilés and Dias (2008), and Bezerra Neto et al. (2002), who cite Tukey's as the most commonly used test in articles published in scientific journals.

The Tukey test is undoubtedly the most widespread in scientific articles, but a discriminated use would be more adequate, since for each set of treatments there is an ideal statistical method; for example, in a data set with treatments in a nonorthogonal contrast where the interest is to compare all pairs of means with each other, one should choose a procedure of multiple mean comparisons, such as the Tukey and Duncan tests (Perecin & Malheiros, 1989; Carmer & Walker 1985; Petersen, 1977).

However, using mean comparison methods for quantitative factors would be inappropriate and induce the wrong conclusions. Therefore, the scientist has to know the type of data set of his study, such as whether the factors are qualitative or quantitative. If the objective is to compare treatments with the control, the Dunnett test should be used. When the number of treatments is large, to eliminate ambiguity, a mean grouping test, such as Scott-Knott, is recommended.

Regression analysis was used in 385 articles, thus accounting for 31.1% of the total articles published in *Acta Scientiarum. Agronomy* (Table 4). Lower results were observed in the *Journal of Fruits* and the *Revista Brasileira de Fruticultura* (Cantuarias-Avilés & Dias, 2008), in which 1.4% and 15.3% of the published articles, respectively, used regression methods. Of the different regression methods, the polynomial was the most frequently used (83.6%) and the other methods in 16.4% of the articles.

Multivariate methods were used in 210 articles, thus representing 17% of the articles published in *Acta Scientiarum. Agronomy*. Among the multivariate methods, mean grouping by Scott-Knott was found in 54.8% of the articles. The clustering or optimization methods, such as Tocher's method and dendrograms based on the Mahalanobis, Ney and Euclidean distances, were used in 17.6% of the articles and were most frequently used in articles dealing with molecular markers.

Despite the low frequency of each multivariate method, this census showed that publications with multivariate methods were on the rise over the years. Similar results were reported by Cantuarias-Avilés and Dias (2008) in the publications of the *Journal of Fruits* and *Revista Brasileira de Fruticultura*, in which more multivariate methods were present in the last years of publication.

The availability of statistical programs, such as SAS (SAS, 2017) and R (R Core Team, 2017), together with an increase in computational power, allow for the use of more complex statistical methods in a shorter amount of time, which may be related to the recent increase of publications with multivariate methods. In areas of biology, physics, sociology, and medical sciences, there is a steady increase in scientific articles using multivariate analysis (Silva, Wanderley, & Santos, 2010).

Nonparametric methods were used in only 36 articles, thus representing 2.9% of the articles published in *Acta Scientiarum. Agronomy*. Among these, the authors mostly used the chi-square (47.2%) and Kruskal-Wallis tests (36.1%). Other methods were used in only 6 articles, thus representing 16.6% of the methods (Table 4). The method of the least square parameter estimation was used in 79.6% of the articles published in *Acta Scientiarum. Agronomy*. However, other methods, such as Bayesian inference and likelihood, were reported in some of the published articles (Table 5).

In the review of the articles published in *Acta Scientiarum. Agronomy*, 107 articles with correlation analysis were identified (Table 5). Of these, approximately 87% used Pearson's parametric correlation, while the other correlations that were observed represented 13% of the articles (Table 5).

Pearson and Spearman correlation analyses are widely used by researchers in the search for relationships between characteristics, although the method does not represent a cause-effect relationship. However, some researchers use simple correlations to explain uncorrelated phenomena or phenomena that do not involve the biological influence of one trait on the other but still exhibit high correlation values. In this context, researchers who seek to solidly explain the cause-effect relationship of two traits should use more accurate statistical methods, such as partial correlations and path analysis.

Table 5. Classification of articles published in *Acta Scientiarum. Agronomy* between 1998 and 2016, classified according to their use of methods for estimation of parameters, correlations and statistical programs.

Method	Number of papers	%
Least squares	985	79.63
Likelihood	1	0.08
Correlation	107	8.65
Pearson (parametric)	93	86.92
Spearman (nonparametric)	9	8.41
Cophenetic correlation	3	2.80
Partial (parametric)	1	0.93
Genotypic correlation	1	0.93
Computer program		
Not mentioned	552	49.46
SAS	111	9.95
GENES	78	6.99
SISVAR	77	6.90
SAEG	62	5.56
SANEST	33	2.96
R	24	2.15
STATISTICA	19	1.70
ESTAT	15	1.34
ASSISTAT	11	0.99
SIGMA PLOT	10	0.90
Others	124	11.11

In this study, 67 different programs were identified in the published articles, thus demonstrating the great diversity of statistical programs. The most commonly used was SAS (10%), followed by GENES (7%), SISVAR (6.9%), SAEG (5.6%), SANEST (3%), R, (2.2%) and STATISTICA (1.7%). In articles published in agrarian, national and international journals, Montanhini Neto and Ostrensky (2013) and Cantuarias-Avilés and Dias (2008) stated SAS as their preferred program. In 49.5% of the published articles, the authors did not mention the statistical program (Table 5). The data corroborates Montanhini Neto and Ostrensky (2013), who found that 49% of the articles that used statistical methods contained no information as to the use of any computer program.

The classification of articles regarding the adequacy of statistical methods is shown in (Table 6). To evaluate the correct use of statistics in the articles, the assumptions were not considered as a penalty. In most of the articles (76.4%), the statistics were adequate to process the data and, consequently, their conclusions were based on the best possible statistical inferences.

Table 6. Numbers and percentages of articles with an appropriate use of statistical methods and of articles in which errors were committed in *Acta Scientiarum. Agronomy* from 1998 to 2016.

Use of statistical methods	Number of papers	%				
Adequate	853	76.43				
Suboptimal	184	16.49				
Inadequate	79	7.08				
Incorrect use of procedure						
Mean comparison	160	60.84				
ANOVA	40	15.21				
Regression	32	12.17				
Descriptive statistics	17	6.46				
Correlation	11	4.18				
Multivariate	3	1.14				
Errors by area						
	Suboptimal (n°)	Suboptimal (%)	Inadequate (n°)	Inadequate (%)	Total (n°)	Total (%)
Crop Production	108	16.11	58	8.65	166	24.77
Crop Protection	37	26.24	13	9.21	50	35.46
Soil Science	12	8.21	8	5.47	20	13.69
Agricultural Engineering	5	5	10	10	15	15
Microbiology	7	28	0	0	7	28
Genetics and Breeding	5	3.85	0	0	5	3.85

The articles with a suboptimal use of statistics (16.5%) applied methods that did not provide the best inferences or used two methods due to the uncertainty about which method to use or to not knowing the best-suited method. An example of this is the use of mean comparison methods when the number of treatments is large or when a mean grouping method would be adequate, thus avoiding ambiguity, separating the groups more clearly and simplifying the conclusions.

In addition, in experiments with quantitative treatments, the researchers frequently use regressions and a method of the comparison or grouping of means. In this case, researchers make two mistakes, first, by using a qualitative data analysis in quantitative treatments (mean tests), and, second, by presenting two discussions and conclusions for the same data set.

The number of papers with inadequate statistics in which statistical methods were not recommended for the data set was low, representing 7% of the articles published in *Acta Scientiarum. Agronomy*. This shows that the peer review of articles is rigorous and that research that is not up to the statistical precepts is turned down.

The percentage of articles with adequate statistics was higher when compared to the *Revista da Sociedade Brasileira de Zootecnia (SBZ)*, in which 24.6% of the mean comparison methods were correct, 11.2% were partially correct and 64.2% were incorrect between 1984 and 89 (Cardelino & Siewerdt, 1992), and when compared to the *Revista de Pesquisa Agropecuária Brasileira (PAB)*, where the mean comparison methods were adequate in 57%, partially adequate in 11.5% and inadequate in 35.5% of the cases between 1980 and 1994 (Santos, Moreira, & Beltrão, 1998).

In a survey of the *Revista Horticultura Brasileira (RHB)* from 1983 to 2000, the authors concluded that 65.6% of the mean comparison methods were adequate, 22.8% were partially adequate and 11.6% were

inadequate (Bezerra Neto et al., 2002). According to Lee (2010), 51% of the articles published in biomedical journals applied the statistical methods erroneously. In the journal of the Archives of Veterinary Science, 32.2% of the statistical methods were adequate, 33.5% were partially adequate and 34.3% were inadequate in the decade from 2000 to 2010 (Montanhini Neto & Ostrensky, 2013).

The most frequent errors committed by authors in publications in *Acta Scientiarum. Agronomy* consisted of the use of mean comparison methods (60.8%) when applied to quantitative data, which is an error because the researcher loses the intermediate data of the treatments. For example, the best dose of a given product may be within the range of tested doses and, with the use of regression curves, the ideal dose can be identified and recommended, even if it was not tested initially in the treatments. Another frequent error was tests with ambiguous results in the case of a high number of treatments.

Errors were observed in relation to the application of ANOVA (15.2%) when the analysis of the treatment effect indicated nonsignificance, but the authors continued to use mean comparison tests. This assessment was based on the concepts of the protected *F* test (Vieira, 2006). Another error was observed in the use of regressions with only three observed points, thus limiting the analysis to a linear regression without investigating other curves such as the quadratic, which might have a better fit and could explain the results of the response variable.

According to Bezerra Neto et al. (2002), the probable causes of errors may be associated with a lack of knowledge about alternative procedures to multiple mean comparison methods, conditions for the adequate use of the methods and about the studied data type, and a lack of ability by the authors to interpret results, thus causing them to draw erroneous conclusions, although they are using methods that they know well.

The problem of errors can be ascribed to researchers' academic formation, since most are instructed about statistical techniques with an emphasis on the mathematical components, whereas the adequacy of the methods or interpretation of results are little or not considered. Another factor that possibly accounts for many errors is that researchers only think about statistical methods after having obtained the data and/or after a journal's reviewer returns the article to correct the statistical analysis (Glickman et al., 2010).

Among the areas observed in the journal *Acta Scientiarum. Agronomy*, crop protection had the highest percentage of errors in the articles (35.46%), 26.24% of which were in suboptimal statistics. The second area with the most errors was microbiology (28%), followed by crop production (24.77%), agricultural engineering (15%), soil science (13.69) and genetics and breeding (3.85%).

The dissemination of the high proportion of articles with errors in the areas should be considered a constructive review and serve for professional reflection. Why are there so many errors? Are readers able to trust the articles? What about the misinterpretation of results prejudicing the research? How can we improve, try to understand statistical methods, and make partnerships with professionals who understand statistics? These are questions that must be answered in order to continuously improve scientific research.

Conclusion

The articles published in *Acta Scientiarum. Agronomy* were mostly contributions from the field of plant production.

Mean comparison methods were the most commonly applied statistical methods, and, consequently, represent the major causes of errors in the publications of the journal.

Acknowledgements

The authors thank to Carlos Alberto Scapim, from Universidade Estadual de Maringá, Mauricio Carlos Kuki, from Universidade Estadual de Maringá, and Rodrigo Ivan Contreras Soto, from Universidad de O'Higgins.

References

- Banzatto, D. A., & Kronka, S. N. (2013). *Experimentação agrícola*. Jaboticabal, SP: Funep.
- Barbin, D. (1993). *Componentes de variância*. Piracicaba, SP: Esalq/USP.
- Beaver, R., Mendenhall, W., & Reinnhmuth, J. (1974). *Statistics for management and economics* (2nd ed.). North Scituate, MA: Duxbury.

- Benjamin, D. J., James, B., Johannesson, M., Nosek, B., Wagenmakers, E., Berk, R., ... Johnson, V. (2017). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6-10. DOI: doi.org/10.1038/s41562-017-0189-z
- Bertoldo, J. G., Coimbra, J. L. M., Guidolin, A. F., Mantovani, A., & Vale, N. M. (2008). Problemas relacionados com o uso de testes de comparação de médias em artigos científicos. *Revista Biotemas*, 21(2), 145-153.
- Bezerra Neto, F., Nunes, G. H. S., & Negreiros, M. Z. (2002). Avaliação de procedimentos de comparações múltiplas em trabalhos publicados na revista Horticultura Brasileira de 1983 a 2000. *Horticultura Brasileira*, 20(1), 5-9.
- Callegari-Jacques, S. M. (2003). *Bioestatística: princípios e aplicações*. Porto Alegre, RS: Artmed.
- Campos, H. (1983). *Estatística experimental não paramétrica*. Piracicaba, SP: Esalq/USP.
- Cantuarias-Avilés, T., & Dias, C. T. S. (2008). Utilização de técnicas estatísticas em duas revistas de fruticultura. *Ciência Rural*, 38(8), 2366-2370.
- Cardellino, R. A., & Siewerdt, F. (1992). Utilização adequada e inadequada dos testes de comparação de médias. *Revista da Sociedade Brasileira de Zootecnia*, 21(6), 985-995.
- Carmer, S. G., & Walker, W. M. (1985). Pairwise multiple comparisons of treatment means in agronomic research. *Journal of Agronomic Education*, 14(1), 19-26.
- Conceição M. J. (2008). Leitura Crítica dos Dados Estatísticos em Trabalhos Científicos. *Revista Brasileira de Anestesiologia*, 58(3), 260-266.
- Ferreira, D. F. (2008). *Estatística multivariada*. Lavras, MG: UFLA.
- Fisher, R. A. (1971). *The design of experiments*. New York, NY: Hafner Publishing Company.
- Glickman, M., Ittenback, R., Nick, T. G., O'Brien, R., Ratcliffe, S. J., & Shults, J. (2010). Statistical consulting with limited resources: applications to practice. *Chance*, 23(4), 35-42.
- Guo, J., Alemayehu, D., & Shao, Y. (2010). Tests for normality based on entropy: divergences. *Statistics in Biopharmaceutical Research*, 2(3), 408-418.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773-795.
- Lee, J. K. (2010). *Statistical bioinformatics: for biomedical and life science researchers*. Hoboken, NJ: Wiley-Blackwell.
- Lopes, S. J., & Storck, L. (1995). A precisão experimental para diferentes manejos na cultura do milho. *Ciência Rural*, 25(1), 49-53.
- Lúcio, A. D., Lopes, S. J., Storck, L., Carpes, R. H., Lieberknecht, D., & Nicola, M. C. (2003). Características experimentais das publicações da ciência rural de 1971 a 2000. *Ciência Rural*, 33(1), 161-164.
- Montanhini Neto, R., & Ostrensky, A. (2013). Assessment of the use of statistical methods in articles published in a journal of veterinary science from 2000 to 2010. *Acta Scientiarum. Technology*, 35(1), 97-102.
- Perecin, D., & Malheiros, E. B. (1989). *Procedimentos para comparações múltiplas*. Lavras, MG: UFLA.
- Petersen, G.R. (1977). Use and misuse of multiple comparison procedures. *Agronomy Journal*, 69(2), 205-208.
- R Core Team. (2017). *R: a language and environment for statistical computing*. Vienna, AU: R Foundation for Statistical Computing.
- Santos, J. W., Moreira, J. A. N., & Beltrão, N. E. M. (1998). Avaliação do emprego dos testes de comparação de médias na revista Pesquisa Agropecuária Brasileira (PAB) de 1980 a 1994. *Pesquisa Agropecuária Brasileira*, 33(3), 225-230.
- Silva, A. C., Wanderley, C. A. N., & Santos, R. (2010). Utilização de ferramentas estatísticas em artigos sobre Contabilidade Financeira – um estudo quantitativo em três congressos realizados no país. *Revista Contemporânea de Contabilidade*, 1(14), 11-28.
- Snedecor, G. W., & Cochran, W. G. (1967). *Statistical methods* (6th ed.). Ames, Iowa: Iowa State University.
- Statistical Analysis System [SAS]. (2017). *User's guide. Statistic*. Cary, NC: SAS Institute Inc.
- Storck, L. (2000). *Experimentação vegetal*. Santa Maria, RS: UFSM.
- Steel, R. G. D., & Torrie, J. H. (1960). *Principles and procedures of statistics*. New York, NY: McGraw Hill Book.
- Vieira, S. (2006). *Análise de variância (ANOVA)*. São Paulo, SP: Atlas.
- White, S. J. (1979). Statistical errors in papers in the British Journal of Psychiatry. *British Journal of Psychiatry*, 135(1), 336-342.