

Point of View**ANALYZING ONE-WAY EXPERIMENTS: A PIECE OF CAKE OR A PAIN IN THE NECK?**

Marcin Kozak

Warsaw University of Life Sciences - Dept. of Experimental Design and Bioinformatics - Nowoursynowska 159, 02-776 Warsaw, Poland. e-mail <nyggus@gmail.com>

**ABSTRACT:** Statistics may be intricate. In practical data analysis many researchers stick to the most common methods, not even trying to find out whether these methods are appropriate for their data and whether other methods might be more useful. In this paper I attempt to show that when analyzing even simple one-way factorial experiments, a lot of issues need to be considered. A classical method to analyze such data is the analysis of variance, quite likely the most often used statistical method in agricultural, biological, ecological and environmental studies. I suspect this is why this method is quite often applied inappropriately: since the method is that common, it does not require too much consideration—this is how some may think. An incorrect analysis may provide false interpretation and conclusions, so one should pay careful attention to which approach to use in the analysis. I do not mean that one should apply difficult or complex statistics; I rather mean that one should apply a correct method that offers what one needs. So, various problems concerned with the analysis of variance and other approaches to analyze such data are discussed in the paper, including checking within-group normality and homocedasticity, analyzing experiments when any of these assumptions is violated, outliers presence, multiple comparison procedures, and other issues.

**Key words:** analysis of variance, assumptions, graphical statistics, multiple comparisons, normal distribution, non-parametric statistics, one-way designs, statistical analysis

**ANÁLISE ESTATÍSTICA DE EXPERIMENTOS COM UM ÚNICO FATOR: MUITO FÁCIL OU MUITO DIFÍCIL?**

**Resumo:** Realizar análises estatísticas pode ser complicado. Em situações práticas muitos pesquisadores utilizam os procedimentos de análise mais comuns, sem investigar se os mesmos são apropriados para os seus resultados, ou mesmo se há outros métodos que poderiam ser mais adequados. Nesse artigo buscarei mostrar que mesmo na análise de experimentos de classificação simples (com um único fator) vários aspectos precisam ser considerados. A forma clássica de análise desse tipo de dados é a análise de variância, que é provavelmente o método estatístico mais usado na agricultura, biologia, ecologia e estudos de meio ambiente. Suspeito que essa é a razão pela qual tal método é frequentemente usado de forma inapropriada: uma vez que ele é muito usado, não suscita maiores considerações. Imagino que seja esse raciocínio que muitos pesquisadores devam empregar. Análises incorretas podem fornecer falsas interpretações e conclusões, e dessa forma é importante prestar atenção na escolha do procedimento a ser usado na análise. Não estou sugerindo que algum método difícil ou complexo deva ser usado, mas sim que um método correto seja adotado, de forma a fornecer os resultados adequados. Dessa forma, vários problemas relacionados à análise de variância e outras abordagens para analisar esse tipo de dados são discutidas nesse artigo, incluindo verificações de normalidade e homogeneidade de variâncias, análise de experimentos com violação dessas pressuposições, presença de dados discrepantes, testes de comparações múltiplas, além de alguns outros problemas.

**Palavras-chave:** análise de variância, pressuposições, análises gráficas, comparações múltiplas, distribuição normal, estatística não paramétrica, experimentos de classificação simples, análise estatística

**INTRODUCTION**

Analysis of variance (ANOVA) is an omnipresent approach towards analyzing various designs in agri-

cultural, biological, ecological and environmental studies (see for example, in the following recent *Scientia Agricola* issue: Caires et al., 2008; Crusciol et al., 2008; Cruz & Cicero, 2008; Cruz et al., 2008; Ferreira

et al., 2008; Guimarães et al., 2008; Miyauchi et al., 2008; Morgado & Willey, 2008; Oviedo et al., 2008; Santos et al., 2008; Soares et al., 2008; Vieira et al., 2008; Yamamoto et al., 2008). Among these designs, of course, is the most basic, one-way design. The analysis of variance, proposed by Ronald Fisher in the first decades of the twentieth century, has been further developed and has evolved in various directions. Yet, the main idea remains the same, and most applications for single-factor data require answering the ANOVA-like question: whether or not the groups differentiate a character of study in terms of the mean.

Nonetheless, data vary and each data set has its own specificity. ANOVA has some assumptions that for many data sets are violated; sometimes these violations can be ignored, sometimes they must not. To apply the method in such a situation, one may need to act somehow so that ANOVA works as a researcher desires. If one merely applies the method for any one-way data without much consideration what actually the data look like, then in all-too-many cases the analysis will be incorrect. "Incorrect" does not only mean incorrect from a statistical point of view; first of all, it means that the answer to the question the researcher asks has big odds to be incorrect. Everything must be done to avoid this undesirable situation.

There are many paths to follow when analyzing such kind of data. These can be found in various sources on statistical analysis for agricultural, biological, ecological and environmental sciences, including Sokal & Rohlf (1995), Quinn & Keough (2002), Gotelli & Ellison (2004) and many others. Because the problem is quite intricate, researchers still try to find the best possible paths. For example, recently Kobayashi et al. (2008) proposed a unified approach towards analyzing one-way toxicity data. I am in general against any unified approaches in data analysis. This is simply because data are very diverse, and one approach will work well for one data set, but may totally fail for another.

In this paper I would like to probe into the issue of analyzing data from one-way experiments. There are many various methods that may be applied to such data, and which is the best will never be decided. This kind of analysis is full of traps and tricks to deal with them. Here I would like to discuss some of these traps and tricks, and to direct the readers' attention to chosen issues of the topic. I will not give any unified approaches, and I will not say what must and what must not be done in a particular situation. Instead, I will just show that even such simple data as those from one-way experiments may be tricky to analyze, let alone data from more complex experiments.

### Checking normality

Kobayashi et al. (2008) point out that toxicity researchers often pay much attention to the normality of a variable within groups; this is not limited to toxicity researchers. Together with the variance homogeneity, the normality is often suggested the most critical when applying ANOVA. This is not necessarily correct in every situation. Gotelli & Ellison (2004, p. 296) say, "Thanks to the Central Limit Theorem..., this assumption is not too restrictive, especially if samples sizes are large and approximately equal among treatments." As Quinn & Keough (2002, p. 192) point out, the  $F$  test is robust to violation of the normality assumption if only data are balanced (which means that sample sizes from groups are the same), the within-group variances are homogeneous (see the following section), and the distributions in groups are not too skewed. If any of these conditions is not met, however, the lack of normality may noticeably affect the results by increasing the actual probability of type I error.

Another problem is how to check whether the variable is normal within groups. Many authors recommend statistical tests; so do Kobayashi et al. (2008), from among several considered recommending Shapiro-Wilk test (Shapiro & Wilk, 1965) as the most "powerful" (in inverted commas because the power considered by the authors seems not to be the power in a statistical sense). Nonetheless, in spite of this great attention paid to tests for normality, many say such tests are not the best way to check the normality for ANOVA purposes. This is for several reasons. First of all, like all statistical tests, they strongly depend on sample size (from groups in our case). Statistical power of such tests increases with an increase in sample size, which is why it may be difficult not to reject a hypothesis with a large sample, and difficult to reject it with a small sample (Quinn & Keough, 2002, p. 192). Paraphrasing the words of Shipley (2002, p. 191), as sample size increases, we run a greater and greater chance of rejecting our hypothesis on normal distribution because of very minor deviations from the normality that might not even interest us. On the other hand, checking for normality makes sense only when sample size is not too small. Of course, one can apply the Shapiro-Wilk test for three-element sample, but practically, the statistical power of such a test is very small. Besides, as mentioned above, the  $F$  test may work well even despite the lack of within-group normality, so not applying the classical ANOVA only because of the modest lack of normality is not necessarily correct when the other assumptions are met.

Kobayashi et al. (2008) compared Kolmogorov-Smirnov, Lilliefors, Shapiro-Wilk and Chi-square (at three different class widths) tests as well as visual examination of the distribution based on normal probability paper. Their results provide yet another interesting proof of why statistical tests for checking normality should be applied with great caution in the present context. Results of two tests (and the decisions based on them) may be very distinct, even if this is just the Chi-square test applied with different widths of class. Therefore, the choice of a particular test may have quite an impact on the conclusions, which gives much room for subjectivity, for certain undesirable.

If not tests, what might be applied to check the normality? There are graphical methods that can be used for distribution checking. Histograms are the best-known tool for graphical presentation of distributions, but they have been criticized by many statisticians; see for example Farnsworth (2000) for a convincing discussion on why histograms are subjective. Quinn & Keough (2002, p. 61) suggest using the boxplot—the ingenious invention of John Tukey (1977)—for several reasons. First, the central value is given by median, so the shape of a distribution is robust to outliers. Second, they detect outliers in a sample (see the section about outliers below). Third, they show whether the variable is symmetrical or not. Reese (2005) puts it, “Make it a rule: never do ANOVA without a boxplot.” Note, however, that boxplots will not detect multimodality in a variable’s distribution; they also require quite a big sample—a boxplot for three observations does not make too much sense. Probability plots are another useful graphical tool (Quinn & Keough, 2002, p. 62) that may help detect deviations from normality such as skewness and multi-modality. See also Cleveland (1993, 1994) for a comprehensive account of graphical methods for checking distributions.

Here the problem of sample size comes out; Gotelli & Ellison (2004, p. 150) gives the “Rule of 10”, according to which for each treatment one should collect at least 10 replications. If the Rule of 10 is met, graphs will help; otherwise, especially with a really small number of replications (3-4, say, quite a common number in agricultural experimentation), then not only graphs but any other method will not work well in checking normality. The Rule of 10 applies not only to checking normality, but as Gotelli & Ellison say, to experimentation in general.

#### **Checking homogeneity of variances**

For ANOVA, the assumption of variance homogeneity is much more important than that of the within-group normality of the variable. Robust for violation

of the normality assumption, the  $F$  test does not work well with heterogeneous within-group variances, which is why some special attention needs to be paid to the analysis when this situation occurs.

First of all, let us note that the above-given comments on hypothesis testing refer to this situation as well. With small samples, to reject the hypothesis that the variances are equal is difficult due to small statistical power; with large samples, the situation is opposite, and to reject such a hypothesis is rather easy due to large statistical power. This should always be remembered when applying ANOVA and checking its assumptions.

How should one act, then, to check if this assumption is valid? Quinn & Keough (2002, sec. 8.3.2), criticizing statistical tests as a tool for preliminary check of this assumption, suggest applying boxplots of observations within groups or plotting the spread of residuals against group means; Cleveland (1993, 1994) also supports graphing residuals by means of various plots. I agree with these suggestions, though I am aware that many would not; this is a somewhat subjective matter, but the arguments against applying tests given by Quinn & Keough (2002) and summarized here seem convincing.

#### **Dealing with non-normal variables and/or heterogeneous within-group variances**

If the variable is non-normal to the extent that it indeed should not be analyzed with ANOVA (for example, due to large skewness) and/or within-group variances are heterogeneous, various approaches may be applied. Let us mention some of them.

First of all, instead of the regular  $F$ -test, one can apply other tests that were constructed to deal with unequal variances; Quinn & Keough (2002, sec. 8.5.1) list Welch’s test and its modification by Wilcox, Brown-Forsythe test, James second order method, and  $Z$  test by Wilcox; they also mention that in particular situations generalized linear modeling may be of help “if the underlying distribution of the observations and residuals is known, and hence why a transformation might be effective.”. Such tests are still investigated and new approaches are proposed (see, e.g., Xu & Wang 2008). Quinn & Keough (2002, p. 196) and Gotelli & Ellison (2004, pp. 109-116) mention also randomization tests.

Transformations also seem a sensible approach. One needs to choose from among various transformations, the most common being the log and root ones. There is a wide Box-Cox family of transformations (Quinn & Keough, 2002, p. 66; refer to Gotelli & Ellison, 2004, pp. 223-235, and Cleveland 1993, pp. 42-67, for a clear and thorough description of various transformations), which includes the two above-

mentioned transformations as well as many others. Thus, because it searches for the optimum transformation factor over a wide range of transformations, there is quite a chance that it will indeed work well. Nonetheless, it does not have to, in which case one may try the rank transformation. It leads to rank-based tests, usually called the non-parametric tests or even non-parametric ANOVA. One such test, likely the most common and well known, is the Kruskal-Wallis test; as Quinn & Keough (2002, p. 196) explain, this test is incorrectly called the non-parametric ANOVA because no partitioning of variance is involved in this test. Weldon (2005) goes even further and correctly claims that such tests as the Kruskal-Wallis are *not* non-parametric because they are based on location parameters such as the median, while non-parametric methods should work outside of the parametric framework of estimation and hypothesis testing.

The other test is indeed the rank-based ANOVA because it consists of transforming the variable to ranks and applying ANOVA for the transformed variable; Quinn & Keough (2002, p. 196) call this approach the rank transform method. As they state, both these tests (the Kruskal-Wallis test and rank-transform method) provide the same results for one-way ANOVA (which design we discuss in this paper), but the latter is more general and can be applied to more complex designs. Note, however, that Gotelli & Ellison (2004, p. 212) argue that although rank statistics are used commonly by some ecologists and environmental scientists, such an approach “wastes information that is present in the original observations.”

One extremely important thing is that if one applies a rank-based method, it does not mean that the variances (of the rank-based variables) may differ among groups. If they do, one compares neither means nor medians, but distributions, so both the central tendency and the distribution’s shape—including variances, of which we know that are heterogeneous. Hence the application of the rank-based tests to compare the central tendency (medians) under the heterogeneity of within-group variances should be avoided. Unfortunately, as Kobayashi et al. (2008) put it, “The data are examined [by most toxicologists] for homogeneity of variance and if the variance is homogeneous, parametric tests are used and for heterogeneous variance non-parametric tests are used”.

### Outliers

One or more outliers may be a likely reason for both within-group non-normal (even highly skewed) distributions of the variable and heterogeneous within-group variances. Identifying outliers may be carried out by applying graphical tools (including those mentioned above)

and many various techniques (see Hodge & Austin, 2004 for a detailed account of outlier detection).

An outlier can be a mistake, in which case it should be either corrected or removed. It can also be what we could call the *true outlier*. The true outlier is a value that is unusual for other reasons than a mistake; it represents some unusual phenomenon which may happen in nature (and happens, since the outlier has been observed in the study). Note that sometimes it is impossible to find out which is the case; for example, when an outlier comes from a single measurement that cannot be checked, it is rather difficult to decide which version is the case. Nonetheless, if this is indeed a true outlier, one can try to somehow analyze the data with the outlier with robust methods (Quinn & Keough, 2002, sec. 4.5), although this may sometimes be difficult—the outlier may still have quite an impact on the analysis. One can also remove the outlier from the data set, but use it somehow in interpretation; the outlier may indicate that some untypical phenomena can happen, which may need further, more detailed studies. Pedhazur (1982, p. 38) puts it in this way: “... extreme residuals may occur in the absence of any errors. In fact, these are the most interesting and intriguing. ... Discovery of such occurrences may lead to greater insights into the phenomenon under study and to the designing of research to further explore and extend such insights.” Rawlings et al. (1998, p. 331) say, “The outlier might be the most informative observation in the study.”

In summary, unless you have convincing reasons and are sure what you are doing, do not simply get rid of outliers, treating them like an evil. One outlier can be more intriguing and interesting than a bunch of regular observations, than all the remaining observations from the experiment.

### Multiple comparison procedures: once more about them

Multiple comparison procedures (or multiple range tests) are considered an intrinsic feature of ANOVA. Once the general null hypothesis is rejected, multiple comparisons are commonly performed to divide the factor levels into homogenous groups, a homogenous group being the group containing the factor levels for which the means do not differ.

Wildly used, multiple comparison procedures are thought of as a natural extension of ANOVA; I suppose most researchers share this opinion. This is very worrying information because these procedures are one of the topics of applied statistics that cause many concerns, misconceptions, misunderstandings and in general, problems. Hence we should devote some space to multiple comparison procedures and discuss it once more.

Why once more? For the very simple reason that so much has been said about why multiple comparisons should not be used whatsoever; for example, see very convincing and compelling discussions in Saville (2003) and Webster (2007). Yet multiple comparisons are used, interpreted and based on by so many scientists around the world. Yet so many journals, even those really good ones, urge their authors to use these procedures. Yet many would argue that multiple comparisons are the best way to interpret differences among treatment means from the analysis of variance model.

In general, multiple comparison procedures were constructed to overcome the problem of increasing the probability of type I error that occurs when more than one hypotheses are verified simultaneously (hence the name ‘simultaneous’ or ‘multiple’ testing)—in pair-wise (or *post hoc*) comparisons every pair of factor levels is compared. The most popular tests for mean separation that are resistant to multiple testing are for example Fisher’s protected LSD, Tukey’s HSD, Scheffé’s, Duncan’s and Neuman-Keul’s tests. Researchers still work on multiple comparison procedures (e.g., Conagin & Barbin., 2006 and Conagin et al., 2008).

Webster (2007) criticizes the common understanding of pair-wise comparisons in ANOVA with the following words: “Investigators who compare every pair of means by one of the above-mentioned tests seem not to appreciate the difference between a whole experiment, for which these techniques have been developed, and individual comparisons of interest.” Saville (2003) calls these tests inconsistent, explaining, “I call a procedure ‘inconsistent’ if the probability of judging two treatments to be different depends on either the number of treatments included in the statistical analysis, or on the values of the treatment means for the remaining treatments”.

So, multiple comparison procedures were constructed to overcome the theoretical problems concerned with multiple testing. Nonetheless, in most, if not all, situations the practical analysis does not call for such an action: investigators are usually interested in testing the difference between two particular treatments irrespective of other treatments. They ask, “Do treatments A and B differ in mean of a character studied, no matter how many treatments are studied in the experiment?”, rather than, “Do treatments A and B differ in mean of a character studied, when I compare (say) four treatments?”. Webster (2007) puts it, “Further, one must ask why the inclusion of more treatments in an experiment should diminish the power of a test to detect true differences, which is what happens if you apply experiment-wise tests: it does not make sense practically”.

Then, the question that suggests itself is whether there is any superiority of analysis of variance in comparing two treatment means from among a number of means over classical *t*-test to compare two means. Of course there is, because the one-way ANOVA model can provide a more precise estimate of the within-group residual variance than the separate analysis of two chosen treatments.

What should then one do instead of applying multiple comparisons? Saville prefers reporting unrestricted LSD (Least Significance Difference), while Webster SED (Standard Error of the Difference); both are more or less equivalent in the sense that the former strictly follows from the latter. The pro of the former is that it is easy to compare two means by checking whether the absolute difference in two treatment means is smaller than the LSD (which means the difference is not significant for the chosen significance level); its con is that it is done for a particular significance level while the reader may want to decide him/herself about the significance level. The SED, on the other hand, does not base on a significance level, and one may simply calculate LSD for the chosen level from the SED; it does require, however, some calculations, and even if these are simple, it does not facilitate reading a paper. Usually it does not really matter which of the two one chooses since they are more or less parallel if only the reader is skillful in interpreting them; authors may decide to provide both ways if only a journal agrees to publish them, and this may be the best mean of presenting such results.

### Yet some other issues

Note that above I have not touched upon all issues concerned with one-way data. For example, there are big differences between observational and experimental studies, both common in environmental and ecological studies. This difference is not necessarily in how the data are analyzed, but rather how the results are interpreted—in the case of observational studies the effect of the experimental factor may be confounded with many other factors, which practically cannot be taken into account. On the one hand, this is a disadvantage of observational studies because we cannot estimate the “pure” effect of the factor of interest. On the other hand, nature plays its own rules, so what is the worth of experiments in which everything is under control? Can such a fully-controlled situation happen in natural conditions? So, both these types of experiments have their pros and cons. Gotelli & Ellison (2004, chapter 6) discuss this topic in detail.

Nonorthogonal data, which are dealt with under unbalanced designs (in which within-group sample sizes differ), are another topic of importance. Although

usually nonorthogonality calls for special treatment in factorial designs, for one-way data it does not cause too many problems. Quinn & Keough (2002, sec. 8.1.6) say that nonorthogonality can be a problem when other assumptions are not met because the *F*-test is less robust to violations of the assumptions for unbalanced designs. Following Underwood (1997), Quinn & Keough (2002) suggest that “experimental and sampling programs in biology with unequal sample sizes should be avoided, at least by design”. This is a sensible recommendation, but in many situations the data will not be balanced. Then one should pay even more attention to checking assumptions to ensure that they are indeed met.

Independency of observations—or rather its lack—is another important topic for the analysis. In a completely randomized design the observations are assumed to be independent. This means that one outcome does not influence another and *vice versa*. For example, if one studies a perennial plant species and measures its traits across several years on the very same plants each year, then these observations are not independent and we have a so-called repeated-measures design. In such a case not taking into account the specific correlation pattern among the observations may cause the results be far from correct. See, for example, Quinn & Keough (2002) and Schabenberger & Pierce (2002) for interesting insights into the problem of repeated-measures analysis.

Yet another issue, quite an important one, is that when a predictor variable is quantitative, it should not normally be treated as a quantitative one, as is often done. Quite often in such instances not analysis of variance, but regression should be applied to analyze the relationship between the dependent variable and the factor. Webster (2007) and Gotelli & Ellison (2004) discuss this topic. The latter authors stress this topic by distinguishing ANOVA designs and regression designs (Gotelli & Ellison, 2004, chapter 7).

Finally, worth mentioning are visual methods of analyzing data. These should not replace the standard statistical inference, but they can greatly enhance understanding the data and their interpretation. Such methods can be applied to one-way data as well as to more complex experiments. Cleveland (1993, 1994) discusses this topic.

## CONCLUSION

Analyzing one-way data may seem simple at first glance, but the truth is quite otherwise. If only any assumption is violated to the extent that the standard *F* test should not be applied, there are many possible paths to follow; the choice of the best path is not easy.

One should follow the general guidelines (such as those given above and in the sources cited) rather than a flow chart (c.f. Kobayashi et al., 2008) which strictly shows what and when one should do; such flow charts are examples of a mechanical, unified approach towards statistical analysis. Mechanical approaches should be avoided in statistical analysis, and so should unify approaches; the key thing in any data analysis is to understand the data. In general, the bottom line is that when indeed any problems with assumptions or generally analysis occur, an investigator should dig deep into the data to find out the best way of analysis.

Before ending, it is worth mentioning that the above discussion is far from being a complete account of designing and analyzing one-way experiments. For example, we did not touch upon (at least directly) such issues as randomization and replication; spatial grain and special extent of a study; confounding factors, which are common in badly designed experiments and in observational studies; fixed and random factors. There are also many possible methods and approaches that have not been mentioned here. The reader can refer to various sources to learn about these aspects; for example, Gotelli & Ellison (2004) present a very nice introduction to such topics, while Quinn & Keough (2002) go into further detail.

Let us recall also that this paper deals with one-way data from completely randomized experiments. Such experiments are not rare (e.g., Barbanti et al., 2007; Beutler et al., 2007; Oreja & González-Andújar, 2007; Revoredo & Melo, 2007; Damin et al., 2008; Silva et al., 2008), and they represent the most basic experimental situation; any more complex design may require more complex methods and approaches. In fact, many of the comments above are not limited to one-way data, but a number of others are. Each design has its own specificity and may need some special treatment. My aim was to direct your attention to the problem that statistics can have many faces. It requires attention and commitment, just as painting requires attention and heart from a painter. If one treats each data set exactly the same way, then the data will hide their most important features. It does pay off, then, to spend some time on the analysis.

I would like to end this paper with the following thought: “If something goes wrong with an analysis, these are not data what fails to be analyzed—it’s the analyst who fails to analyze the data”.

## REFERENCES

- BARBANTI, L.; MONTI, A.; VENTURI, G. Nitrogen dynamics and fertilizer use efficiency in leaves of different ages of sugar beet (*Beta vulgaris*) at variable water regimes. **Annals of Applied Biology**, v.150, p.197-205, 2007.

- BEUTLER, A.N.; CENTURION, J.C.; CENTURION, M.A.P.C.; FREDDI, O.S.; SOUSA NETO, E.L.; LEONEL, C.L.; SILVA, A.P. Traffic soil compaction of an Oxisol related to soybean development and yield. *Scientia Agricola*, v.64, p.608-615, 2007.
- CAIRES, E.F.; BARTH, G.; GARBUJO, F.J.; CHURKA, S. Soil acidity, liming and soybean performance under no-till. *Scientia Agricola*, v.65, p.532-540, 2008.
- CLEVELAND, W.S. **Visualizing data**. Summit: Hobart, 1993, 360p.
- CLEVELAND, W.S. **The elements of graphing data**. 2 ed. Summit: Hobart, 1994, 297p.
- CONAGIN, A.; BARBIN, D. Bonferroni's and Sidak's modified tests. *Scientia Agricola*, v.63, p.70-76, 2006.
- CONAGIN, A.; BARBIN, D.; DEMÉTRIO, C.G.B. Modifications for the Tukey test procedure and evaluation of the power and efficiency of multiple comparison procedures. *Scientia Agricola*, v.65, p.428-432, 2008.
- CRUSCIOL, C.A.C.; ARF, O.; SORATTO, R.P.; MATEUS, G.P. Grain quality of upland rice cultivars in response to cropping systems in the Brazilian tropical savanna. *Scientia Agricola*, v.65, p.468-473, 2008.
- CRUZ, E.D.; CICERO, S.M. Sensitivity of seed to desiccation in cupuassu (*Theobroma grandiflorum* (Willd. ex Spreng.) K. Schum. – sterculiaceae). *Scientia Agricola*, v.65, p.557-560, 2008.
- CRUZ, R.P.; MILACH, S.C.K.; FEDERIZZI, L.C. Inheritance of pinacle exertion in rice. *Scientia Agricola*, v.65, p.502-507, 2008.
- DAMIN, V.; FRANCO, H.C.J.; MORAES, M.F.; FRANCO, A.; TRIVELIN, P.C.O. Nitrogen loss in *Brachiaria decumbens* after application of glyphosate or glufosinate-ammonium. *Scientia Agricola*, v.65, p.402-407, 2008.
- FARNSWORTH, D.L. The case against histograms. **Teaching Statistics**, v.22, p.81-85, 2000.
- FERREIRA, M.D.; SARGENT, S.A.; BRECHT, J.K.; CHANDLER, C.K. Strawberry fruit resistance to simulated handling. *Scientia Agricola*, v.65, p.490-495, 2008.
- GOTELLI, N.J.; ELLISON, A.M. **A primer of ecological statistics**. Sunderland: Sinauer, 2004, 492p.
- GUIMARÃES, A.P.; MORAIS, R.F.; URQUIAGA, S.; BODDEY, R.M.; ALVES, B.J.R. *Bradyrhizobium* strain and the 15N natural abundance quantification of biological N<sub>2</sub> fixation in soybean. *Scientia Agricola*, v.65, p.516-524, 2008.
- HODGE, V.; AUSTIN, J. A survey of outlier detection methodologies. **Artificial Intelligence Review**, v.22, p.85-126, 2004.
- KOBAYASHI, K.; PILLAI, K.S.; SAKURATANI, Y.; SUZUKI, M.; JIE, W. Do we need to examine the quantitative data obtained from toxicity studies for both normality and homogeneity of variance? **Journal of Environmental Biology**, v.29, p.47-52, 2008.
- MIYAUCHI, M.Y.H.; LIMA, D.S.; NOGUEIRA, M.A.; LOVATO, G.M.; MURATE, L.S.; CRUZ, M.F.; FERREIRA, J.M.; ZANGARO, W.; ANDRADE, G. Interactions between diazotrophic bacteria and mycorrhizal fungus in maize genotypes. *Scientia Agricola*, v.65, p.525-531, 2008.
- MORGADO, L.B.; WILLEY, R.W. Optimum plant population for maize-bean intercropping system in the Brazilian semi-arid region. *Scientia Agricola*, v.65, p.474-480, 2008.
- PEDHAZUR, E.J. **Multiple regression in behavioral research: explanation and prediction**. 2 ed., New York: Holt, Rinehart and Winston, 1982, 882p.
- OVIDO, V.R.S.; GODOY, A.R.; CARDOSO, A.I.I. Performance of advanced generation from a hybrid Japanese cucumber. *Scientia Agricola*, v.65, p.553-556, 2008.
- OREJA, F.H.; GONZÁLEZ-ANDÚJAR, J.L. Modelling competition between large crabgrass and glyphosate-resistant soybean in the Rolling Pampas of Argentina. **Communications in Biometry and Crop Science**, v.2, p.62-67, 2007.
- QUINN, G.P.; KEOUGH, M.J. **Experimental design and data analysis for biologists**. Cambridge: Cambridge University Press, 2002. 738p.
- RAWLINGS, J.O.; PANTULA, S.G.; DICKEY, D.A. **Applied regression analysis: a research tool**. New York: Springer Verlag, 1998.
- REESE, A. Boxplots. **Significance**, v.2, p.134-135, 2005.
- REVOREDO, M.D.; MELO, W.J. Enzyme activity and microbial biomass in an Oxisol amended with sewage sludge contaminated with nickel. *Scientia Agricola*, v.64, p.61-67, 2007.
- SANTOS, C.C.; DELGADO, E.F.; MENTEN, J.F.M.; PEDREIRA, A.C.M.; CASTILLO, C.J.C.; MOURÃO, G.B.; BROSSI, C.; SILVA, I.J.O. Sarcoplasmic and myofibrillar protein changes caused by acute heat stress in broiler chicken. *Scientia Agricola*, v.65, p.453-458, 2008.
- SAVILLE, D.J. Basic statistics and the inconsistency of multiple comparison procedures. **Canadian Journal of Experimental Psychology**, v.57, p.167-175, 2003.
- SCHABENBERGER, O.; PIERCE, F.J. **Contemporary statistical models for the plant and soil sciences**. Boca Raton: CRC Press, 2002. 537p.
- SHAPIRO, S.S.; WILK, M.B. An analysis of variance test for normality (complete samples). **Biometrika**, v.52, p.591-611, 1965.
- SHIPLEY, B. **Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference**. Cambridge: Cambridge University Press, 2002, 332p.
- SILVA, D.H.; ROSSI, M.L.; BOARETTO, A.E.; NOGUEIRA, N.L.; MURAOKA, T. Boron affects the growth and ultrastructure of castor bean plants. *Scientia Agricola*, v.65, p.659-664, 2008.
- SOARES, D.M.; GALVÃO, L.S.; FORMAGGIO, A.R. Crop area estimate from original and simulated spatial resolution data and landscape metrics. *Scientia Agricola*, v.65, p.459-467, 2008.
- SOKAL, R.R.; ROHLF, F.J. **Biometry: the principles and practice of statistics in biological research**. 3 ed., New York: W.H. Freeman, 1995, 887p.
- TUKEY, J.W. **Exploratory data analysis**: Reading: Addison-Wesley, 1977, 688p. Addison-(Wesley Series in Behavioral Science; Quantitative Methods).
- UNDERWOOD, A.J. **Experiments in ecology: their logical design and interpretation using analysis of variance**. Cambridge: Cambridge University Press, 1997, 504p.
- VIEIRA, R.D.; TEKRONY, D.M.; EGLI, D.B.; BRUENNING, W.P.; PANOBIANCO, M. Temperature during soybean seed storage and the amount of electrolytes of soaked seeds solution. *Scientia Agricola*, v.65, p.496-501, 2008.
- WEBSTER, R. Analysis of variance, inference, multiple comparisons and sampling effects in soil research. **European Journal of Soil Science**, v.58, p.74-82, 2007.
- WELDON, K.L. Less parametric methods in statistics. **Metodološki Zvezki**, v.2, p.95-108, 2005.
- XU, L.W.; WANG S.G. A new generalized *p*-value for ANOVA under heteroscedasticity. **Statistics and Probability Letters**, v.78, p.963-969, 2008.
- YAMAMOTO, P.Y.; COLOMBO, C.A.; FILHO, J.A.A.; LOURENÇÃO, A.L.; MARQUES, M.O.M.; MORAIS, G.D.S.; CHIORATO, A.F.; MARTINS, A.L.M.; SIQUEIRA, W.J. Performance of ginger grass (*Lippia alba*) for traits related to the production of essential oil. *Scientia Agricola*, v.65, p.481-489, 2008.

Received March 16, 2009

Accepted May 22, 2009