

Best Practices for Presenting Statistical Information in a Research Article

Matthew H. Kramer¹

U.S. Department of Agriculture, Agricultural Research Service, Statistics Group, Building 003, Room 330, 10300 Baltimore Avenue, Beltsville, MD 20705

Ellen T. Paparozzi

Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583-0724

Walter W. Stroup

Department of Statistics, University of Nebraska, Lincoln, NE 68583-0963

Additional index words. author's guidelines, experiment design, data collection, statistical analysis, statistical recommendations

Abstract. A key characteristic of scientific research is that the entire experiment (or series of experiments), including the data analyses, is reproducible. This aspect of science is increasingly emphasized. The Materials and Methods section of a scientific paper typically contains the necessary information for the research to be replicated and expanded on by other scientists. Important components are descriptions of the study design, data collection, and statistical analysis of those data, including the software used. In the Results section, statistical analyses are presented; these are usually best absorbed from figures. Model parameter estimates (including variances) and effect sizes should also be included in this section, not just results of significance tests, because they are needed for subsequent power and meta-analyses. In this article, we give key components to include in the descriptions of study design and analysis, and discuss data interpretation and presentation with examples from the horticultural sciences.

This article provides recommendations for statistical reporting in a research journal article. Appropriate and informative reporting, and the wise use of statistical design and analysis throughout the research process, are both essential to good science; neither can happen without the other. In addition, many journals now require access to original data and the code used for analyses. This article is *not* a statistics tutorial; we do not explain how to do any of the statistical methods mentioned. There are many, many papers and books that provide that information; some are cited in our reference and selected reading section. Instead, we give guidelines for horticultural scientists on how best to incorporate and present statistical information in a scientific paper. We also focus on experimental rather than observational studies. To do the latter justice would require greatly expanding this article, and the majority of papers published by the American Society for Horticultural Scientists are experimental studies. A very useful complementary article is by Onofri et al. (2010), which gives specific advice for many issues we treat only generally.

This paper is divided into two sections, as follows:

Section 1. When Are Statistics Needed and What Is the Purpose of Statistics in a Research Paper?

Section 2. Recommendations for Writing about Statistics in a Research Paper

- What Goes in the Materials and Methods Section?
- What Goes in the Results Section?
- Additional Details and Descriptions about Design, Data Collection, and Analysis
- Pointers for Writing about Statistics for the Horticultural Sciences
- Literature Cited and Selected References

Section 1: When Are Statistics Needed and What Is the Purpose of Statistics in a Research Paper?

The scope of horticultural research is large and not all studies require statistics. For example, anatomical and morphological studies can be purely descriptive. With that said, these kinds of descriptive studies are a subset of observational studies, which also include studies at the genomic, ecologic, and landscape level. For observational studies, there are useful methods for determining associations, clusters, and dimension reduction, to name a few, that are statistics based. In this article we focus primarily on research

questions that require *inferential* statistics. Typically, using designed experiments when addressing a research question requires experiment planning, data collection, and subsequent statistical analysis, and the following recommendations are applicable.

The statistical section in an article serves five general functions. First, the design, data collection, method of analysis, and software used must be described with sufficient clarity to demonstrate that the study is capable of addressing the primary objectives of the research. When adequate information is provided, it allows for an informed peer review and for readers, in principle, to reproduce the study, *including the data analysis*. Second, authors must provide sufficient documentation to create confidence that the data have been analyzed appropriately. This includes verifying required statistical assumptions and justifying choices—such as the chosen mean comparison procedure and any other method that might affect results and conclusions, such as controlling for experimental-wise error. Experiment-wise error rate (or family-wise error rate, depending on how family is defined) is the probability of committing at least one Type I error throughout the whole experiment. Although the error rate for an individual hypothesis test may be small, if one tests many hypotheses, one becomes more likely to declare false significance for at least one. If the tests are not independent (e.g., using the same plants to test multiple attributes or over time, as is common in this field), this can increase the experiment-wise error rate. For example, if a plant in one treatment group is diseased, this will affect all the (correlated) measures of that group, and thus all hypotheses tests. Third, data and their analyses must be presented coherently. The statistical model and analysis should naturally follow from the study design, and be consistent with relevant characteristics of the data, such as the underlying sampling distribution (e.g., normal, Poisson, binomial). Figures and tables should illustrate, and be consistent with, important results from the analysis. Fourth, readers should not have to guess which scientific questions the analysis answers. Effects deemed statistically significant must also be shown to be biologically/economically important. Effects of potential biological/economic importance but whose statistical significance is not supported by the data should also be reported. There is an implicit assumption of adequate power when discussing results from any statistical tests. Power is estimated during the design phase using results from previous experiments or parameter estimates from the literature. Fifth, readers should be able to use information in the statistical reporting section as a resource for planning future experiments. Variance estimates are especially important for this function.

The goal of this article is to provide an overview of how best to communicate statistics used in horticultural research. Therefore, it does not include specifics to address every

Received for publication 5 Feb. 2019. Accepted for publication 2 Apr. 2019.

We thank the reviewers for their excellent comments and reference recommendations.

¹Corresponding author. E-mail: matt.kramer@ars.usda.gov.

contingency. Statistical methods continuously change, with new methods developed to address advances in biologic and ecologic research. For many studies, traditional and familiar methods (a.k.a. “standard statistics”) are adequate. However, for other studies, newer, less familiar methods are preferable, if not essential. Use of newer methods should not be an obstacle for publication.

Section 2: Recommendations for Writing about Statistics in a Research Paper

The following sections outline key points that should be addressed in the Materials and Methods section, and in the Results section of a journal article. Kramer et al. (2016) document common statistical problems for a sample of horticultural articles and should be used as a checklist of mistakes to avoid. The work by Reinhart (2015) is not overly technical and it explains many of these issues and other mistakes further, mostly in a biologic context.

What goes in the Materials and Methods section?

Broadly speaking, there are two main statistical areas that the Materials and Methods section should address: 1) how was the study designed and 2) how were the data analyzed. Recommendations are grouped by subtopic.

Design and data collection.

- The main idea of this section is to provide all information relevant to subsequent statistical analysis and interpretation about the design—specifically, how the experiment was conducted, how the data were collected and subsequently handled up to the point when the data were ready for statistical analysis. These are detailed next.
- Describe the design. There are two components of experimental design: the *experiment* design and the *treatment* design. Both must be described.
- The *treatment* design refers to the organization of treatment factors. Factorial designs (e.g., varieties × potting substrate) and dose–response (e.g., amount of nutrient applied) are familiar examples.
- The *experiment* design refers to how the experimental units were organized and how randomization was done. Familiar examples are the completely randomized design (CRD) and randomized complete block design (RCBD). Any restrictions on randomization (e.g., blocking) or other ways observations were grouped must be described; this is part of the experiment design.
- Describe covariates, if any. Provide the units of replication (the experimental unit; in other words, the smallest unit to which treatments were assigned independently) and the units of observation (sampling unit). The units of replication may differ for different factors (as they do, for example, in a split-plot design).

- Describe how data were collected and how samples were pooled/batched, if this was done. Identify whether these were one-time measurements, multiple measurements on the plant/plot at the same time, repeated measures over time, or measurements on different plant characteristics.
- Provide numbers, so it is clear how many units were in each block/group, how many received each treatment, and so on. Total sample size must be easily calculated, if not given. If a power analysis was used to determine the sample size, provide details. If not, explain how the sample size was determined. For example, one could write: “Growth chambers were limited to 30 plants, and three growth chambers were available. Previous studies using a similar setup and similar plant numbers had no difficulty detecting even moderate differences in growth patterns.”
- Identify which variables are dependent (i.e., the response variables one measures, such as yield, biomass, time to flowering, elemental concentration) and which are independent (see the previous description of treatment design).
- Describe any transformation of variables (e.g., logarithmic transformation) and the reason it was needed; this applies to both dependent and independent variables. Often, dependent variables can be fit without transformation if the appropriate sampling distribution is specified in a generalized linear model. When this is possible, generalized linear models are preferable to variance stabilizing transformations.

Data analysis. Broadly speaking, data analysis includes the following steps:

1. Plot the original data to visualize what has happened in terms of treatment effects, distribution of data, and other features of the data deemed to be important.
2. Determine a statistical model consistent with the study design and the distribution of the data, and mean comparison procedures needed to address the objective of the research.
3. Determine the statistical assumptions associated with the selected model.
4. Select the software to be used to implement the analysis.
5. Run the analysis and verify that the assumptions are satisfied.
6. Report in the Materials and Methods section how the previous steps were completed.
7. Report the outcome of the analysis in the Results section.

What goes in the Results section?

There is no one-size-fits-all way of presenting the results of a statistical analysis. This is true for many aspects of using statistics in horticultural science, making it impossible to give advice covering every situation; instead, we provide general guidelines. Authors must decide what best tells the story of their research results. Tables and

figures are common methods of presenting data results. The following are principles to follow:

- If you include graphics showing the data, presenting data summaries, or depicting results from modeling, the intent is to portray the findings of the research accurately and make it easier for readers to visually understand the data, estimates and findings from the analysis.
- Statistics that appear in both figures and tables should be consistent with the way the data were analyzed. If objectives are addressed using descriptive statistics, then these should appear in a figure or table, along with their appropriate measures of variability.
- If the objectives are addressed using a statistical model, as is usually the case, then statistics obtained from the model should appear in the figure or table, along with their appropriate measures of variability.
- For modeling results and hypothesis testing, there are two main categories of output from statistical software that should be presented: 1) diagnostic information demonstrating that the method and statistical model used are appropriate and 2) parameter estimates and hypothesis tests that bear directly on the research objectives. The connection to the research objectives must be clear for each statistical result (do not simply copy results produced by software). Two other categories of statistical results should be considered: 1) estimates of quantities from the model that may be useful in future research (e.g., variance estimates) and 2) statistical support for unexpected findings.
- Demonstrate that model assumptions were satisfied (this could be just a sentence for simple models). See the previous point.
- For multiple dependent variables, give the correlations among these variables [and possibly the correlations separately for each treatment if the treatments affect the correlations (discussed later)]. Experiment-wide error control may be necessary.

Additional details and descriptions about design, data collection, and analysis

Statistics for the Materials and Methods section. The Materials and Methods section should address the first function given in Section 1. The design, data collection, method of analysis, and software used must be described clearly. When choices were made or when nonstandard procedures were used must be justified.

Description of the study design. This means “design” as broadly defined. If data were collected, whether from an observational study, a survey, or a designed experiment, there was a design. At a minimum, all designs include three elements: The first is the response variable (i.e., the outcome or outcomes measured), the second is the treatment design (i.e., the treatments or conditions

being evaluated), and the third is the design structure of the experiment, which includes the units of replication (called the experimental unit in designed experiments), the units of observation (called the sampling unit in designed experiments), and grouping of units, if any. Grouping may consist of blocking, research conducted at multiple locations, or data collected on multiple occasions.

The following are three scenarios to illustrate these points. *Scenario 1*: Suppose there are plants in flats on a bench. If treatments are assigned randomly and applied to the bench, the bench is the experimental unit. If observations are made on the flat, then the flat is the unit of observation (sampling unit). This is a CRD. *Scenario 2*: If treatments are assigned randomly to individual flats within each bench, then flat is the experimental unit. Bench is a blocking factor. If observations are made on the flat, then the flat is the unit of observation. Notice that the experimental unit and the sampling unit can be identical. This is not the case in scenario 1. This is an RCBD. *Scenario 3*: Experiments with factorial treatment designs often have different-size experimental units for different factors. In this scenario, irrigation or nutrients are applied using drip lines across a bench, but each bench has several flats, with a different variety in each flat. Here, bench is the experimental unit with respect to irrigation/nutrient and flat is the experimental unit with respect to variety. In design language, this is a split-plot experiment, with the bench as the whole-plot experimental unit, irrigation/nutrient is the whole-plot treatment factor, flat is the split-plot experimental unit, and variety is the split-plot treatment factor. See Onofri et al. (2010) for another good example illustrating true and pseudo-replication.

Important note: Although it is acceptable to name the design, such as an RCBD or Latin square design, a name alone is insufficient and may be misleading. So regardless of whether a design name is used, authors must give the treatment factors, the experimental units, sampling units, and the blocking criteria (if any). For example, an RCBD may or may not have treatments replicated in each block. If treatments are replicated, one can test whether a treatment effect is the same in all blocks; if not, one has to assume it is. So, “RCBD” does not contain all the necessary information about the design.

Data collection. This means list the response variables measured and describe how each was measured. It is also beneficial to make various plots of the original data to determine if there is a treatment effect (these plots are not necessarily included in the published paper). The biology should lead the statistics. Beyond this, you are looking for two things. When you describe the response variable, you want to focus on the sampling distribution of the response variable because this affects the model selected for the analysis of the data. You should plot the response variable against the predictor variables and

look for recognizable patterns—in particular, to determine if (and how) variability changes systematically with the mean. For example, these may be scatterplots or boxplots. Another useful plot groups observations in a natural way (say, by treatment combination) and plots the means of the groups against their standard deviations. Many statistical methods assume the response variable is normally distributed, in which case variability should be roughly the same throughout the range of the response variable. A histogram of the *residuals* from the appropriate model with a normally distributed response variable results in a bell-shaped distribution. Note that a histogram of the raw response variable should not have a bell-shaped distribution because, if there really are treatment effects, the histogram should have a peak at each treatment mean.

Many commonly measured response variables in horticulture have a non-normal distribution. For example, germination rate (number of seeds germinated successfully/the number planted) has a binomial distribution. Many variables are continuous but have strongly right-skewed distributions, such as berry weight. A log-normal distribution often works well for this response variable. Generalized linear models allow the data to arise from many processes; the normal distribution is just one of several. Others include the log-normal, gamma, exponential, beta, binomial, Poisson, and negative binomial. The latter three are used to model count data. Again, plots used to assess the data and suggest models are part of your toolbox for determining the formal statistical analysis you will conduct, but usually are not included in an article.

The second thing you are looking for is any aspect of the data collection process that might affect the structure of the experiment design. Milliken and Johnson (2009) give examples in which the data collection process alters the study design. In one example, plants were grown in multiple distinct blocks, but then material for each treatment was combined from all blocks to allow measurement of the micronutrients of interest. The original blocks were legitimate replicates, but combining material precludes estimating block-to-block variability, effectively creating an unreplicated experiment. For this reason, a clear description of the data collection process is essential.

Model description. Model description consists of giving the assumed distribution of the response variable and the sources of variation in the treatment and experiment design.

- Scenario 1: plants assigned to benches in a CRD. The model would simply be Response = Treatment + Experimental error. (Plant-to-plant variability should be the largest contributor to the experimental error component.)

- Scenario 2: treatments assigned to flats in an RCBD, with benches as the blocking criteria. The model would be Response = Treatment + Benches + Experimental error. This model assumes the treatment effect does not differ from bench to bench.
- Scenario 3: Irrigation is the whole-plot treatment factor, benches are the whole-plot experimental units, variety is the split-plot treatment factor, and flat is the split-plot experimental unit. The model is Response = Irrigation treatment + Whole-plot error + Variety + Irrigation × Variety + Split-plot error. This model assumes the irrigation effect does not differ from bench to bench and that the variety effect does not differ from flat to flat. [In statistical jargon, there is no interaction between any of the fixed effects (irrigation and variety) and any of the random effects (bench and flat)].

Other aspects of analysis. Because of the wide range of research subject matter and scales (laboratory to field), we give general principles. First, the statistical software used to analyze the data is *not* the method of analysis. Authors must first describe clearly the statistical procedures to compare or otherwise characterize the treatments. As illustrated in the three previous scenarios, the method of analysis must be consistent with the study design and data collection process. If there are assumptions critical to the validity of the method of analysis used, authors must state that the assumptions were met and how those assumptions were verified. If it is unclear what the assumptions are or how to verify them, talk to a statistician. Third, there must be a clear connection between the statistical methods used and the primary objectives of the research. This is where treatment design comes in, and it is important to match how you compare the treatments with the treatment design. For example, if you are comparing different varieties, then a mean comparison test is appropriate. Depending on the relative seriousness of Type I (false positives) and Type II (false negatives) errors, there are different ways to implement a means comparison test. At one extreme are two tests: Duncan multiple ranges test and an unprotected least significant difference test, neither of which control Type I error. At the other extreme are Scheffé and Bonferroni tests, which offer extreme control of Type I error at the expense of Type II error. There is a time and place for each test. Authors must state which procedure was used and why that procedure was chosen. The treatment design for experiments yielding genomic data is often simple, but the analyses are complicated. When analyzing RNAseq and similar genomic data, controlling for false discovery rate (which is also a multiple-comparisons issue) is similarly important.

In addition to factorial treatment designs [when main effects (factors with discrete levels) and their interactions are important], regression (when one or more predictor variables are continuous) is often used in horticulture. In some cases, continuous predictor

variables are observational in nature. They are often called covariates in designs that also have factors. The distribution of the response variable needs to be stated because that distribution, in part, determines which statistical model is appropriate.

When the assumptions underlying a parametric method are violated, “nonparametric” methods should be used. These are not assumption-free; one assumption is that the response variable has the same sampling distribution across treatments (e.g., always skewed to the right).

Ratios constructed of two random variables (e.g., root mass/aboveground mass) have poor statistical properties (the assumptions of a parametric test are often violated because the variance of the ratios is not well determined). If ratios need to be used in an analysis, consider obtaining advice from a statistician familiar with the analysis of ratio data.

The trend in biological, medical, and social sciences journals is also to report effect sizes rather than simply the results of a significance test [see Nakagawa and Cuthill (2007) for a readable justification and concrete suggestions]. This is now required in many journals (Tressoldi et al., 2013).

With software improvements, Bayesian statistical methodology is gaining acceptance among biologists. In certain cases, such as models with layers of random effects, Bayesian methods enable analyses that would otherwise not be possible. In simpler models, there is often not much difference between results from Bayesian and frequentist (“traditional”) statistical analyses unless there is relevant prior information that improves the accuracy and precision of parameter estimates. Findings based on the use of Bayesian methodology are, in principle, acceptable in most biological journals, although require more explanation for readers to understand the results.

It may not be clear at the onset of an analysis which statistical methodology should be used, and several different kinds of analyses may be done with the same data set to determine which one makes the most sense. For example, diagnostics following fitting a model may suggest that the assumptions are not met. Alternative models may be examined to determine whether they fit the data better. This is not a free pass to try models until one finds the results one desires. Rather, one oscillates between fitting models and judging them using diagnostics until one is satisfied that one has selected a model that both captures the essential features of the data and has its assumptions satisfied. A useful discussion on obvious and not-so-obvious biases resulting from such a path is given by Gelman and Loken (2014). Note that if two reasonable statistical models give contradictory conclusions, authors could present both, as long as sufficient information for the reviewers and readers to understand the issue is provided.

Statistical software. After authors have described the method of analysis, following

guidelines given previously, then any software used for statistical analyses should be cited, including online software. Include the version (the release) in the citation. Software developed by the authors for the analysis and, thus, not generally available should be explained sufficiently (perhaps in an appendix) for readers to understand what it does and why off-the-shelf software was not suitable. Authors must make the software available for others to use upon request and should include well-documented copies of the code for the reviewers. If the software was part of a system, such as SAS® or R, authors must also give the specific procedure used, such as SAS PROC GLIMMIX or the lme4 package in R.

Statistics for the Results section. As with the method of analysis, there is no one-size-fits-all rule for presentation of data and associated formal statistical analysis. Again, we provide general principles.

First, data should be presented so that the relevant information with regard to the study’s primary objectives and most important findings are clear. Presentation may be via figures or tables, as long as these figures or tables inform rather than inadvertently hide or distort important information. In general, a picture is worth a thousand numbers. Well-conceived figures tend to portray the data’s important messages more understandably than tables.

If multiple responses are measured on the same sampling unit, such as weight, height, sugar content, and macro- and micronutrient content in a plant, correlation among these variables is likely and should be accounted for in the analysis (this is a kind of repeated-measures design) and correlation coefficients should be provided. Note that these correlations may change with different treatments or environments, just as mean responses may, so a single set of correlation coefficients may not summarize adequately the relationships among the variables in the experiment. If multiple responses are measured, experiment-wise error control may be needed. The same considerations for balancing Type I and Type II error rates could be applied here, as mentioned earlier.

Anytime means are compared, the *standard error of the difference* must be reported. In most cases, the standard error of a mean can be considered optional. This is admittedly a break with tradition, but it is an essential one. A plot depicting means with standard error bars is, by itself, insufficient.

Formal statistics. Formal statistics include results of hypothesis tests (e.g., *F* or *t* statistics, *P* values), results of mean separation tests, estimates of means, differences, regression coefficients and their associated standard errors or confidence intervals, predicted values and their associated prediction intervals, and so on. In general, providing the mean (or mean difference) and its confidence interval is preferable to reporting only the results of a hypothesis test. Formal statistics should accompany and provide support for, but not substitute for, the depictions of the data described earlier. The American Statistical Association issued a

policy statement in 2016 (Wasserstein and Lazar, 2016) that clarifies legitimate vs. illegitimate uses and interpretations of *P* values associated with hypothesis tests. *P* values tell us whether the observed differences in the data are likely the result of chance or whether there is strong evidence of a true difference. They cannot tell us whether the difference is big enough to matter.

The main message should be that the observed difference is biologically, economically, or scientifically consequential, not that a *P* value was statistically significant. If the treatment group differs significantly from the control group, the emphasis should be on the biological consequences of finding a difference of that magnitude. If a regression line has a significant slope, the emphasis should be on the functional relationship between the independent and dependent variables. What underlying biological principle is responsible for a slope of this size? Let biology lead and let significance tests follow.

Often, not finding a statistically significant difference is important and should be reported if there was sufficient power to detect a biologically important difference. For example, if a study is done on the assumption (perhaps based on conventional wisdom or a previous research report) that a treatment difference exists, and data from a new study suggest otherwise, that information should be reported. Journals do science a major disservice by preferentially reporting only statistically significant results. This practice is called “publication bias” and is increasingly recognized to be a serious issue in all sciences. Sometimes a nondifference is the most important finding.

Pointers for Writing about Statistics for the Horticultural Sciences

Many terms have technical meanings in statistics, as well as more general—and less precise—uses in common language. For example, “significant” has a specific definition in hypothesis testing, but the words “significant” and “important” tend to be used loosely and interchangeably when describing scientific results. It is best to avoid ambiguities in your writing (What is the meaning of “significant findings?”) Instead, describe the difference. For example, for a dry weight measurement, treatment A resulted in a heavier plant than treatment B. Commonly used statistical terms (e.g., analysis of variance) do not need to be defined in the article. Less common ones (e.g., reliability) do need accompanying definitions. If a reference needs to be given for a statistical technique, refer to an easily available (and commonly used) textbook if possible. The second choice would be an article in a horticulture or other biological journal. The third choice is a review article that explains the technique and perhaps compares it with others. The last choice is an article in the statistical literature that requires an advanced background in statistical theory.

Readers of an article may have a different reason for looking at results than the author’s stated purpose (e.g., to compare some of the

results in the article with data readers have from a location they use, rather than the within-location comparison of cultivars in the article), which is another reason why summary information about the original data (e.g., means and standard deviations) needs to be provided. Data summaries may also be used in a subsequent meta-analysis; these typically use means, standard deviations, and other estimated parameters (e.g., block-to-block variance).

Statistics, and figures and tables. Scientific publications are replete with tables, figures, and plots that are easy to read, technically impressive, pretty to look at, but, unfortunately, can be misleading in their content with respect to the objectives of the research they are intended to portray. If a figure shows the results of statistical modeling (e.g., means and their standard errors), you should try including the original data in the figure, perhaps in the background. This helps readers assess the adequacy of the statistical model visually. Rather than reiterate the advice of others, we suggest an excellent source for describing how data (and legends) should be presented: *How to Report Statistics in Medicine* (Lang and Secic, 2006; pp. 325–393).

Plant scientists are not expected to know everything when conducting research, and this is becoming more evident with increasing collaborations across fields of study. Plant scientists should know, however, when they need input from a statistician. If so, we advise meeting with a statistician *before* setting up the experiment. A statistician will not be able to help after data from a poorly designed experiment are collected (other than

to suggest rerunning the experiment with a better design).

A well-designed experiment can often be analyzed a number of ways and, usually, there are choices to make along the way. Examples include whether there is overdispersion, whether interaction terms are necessary, or whether a multivariate analysis should be considered to account for correlation among response variables. Should the statistician be extensively involved in the design and analysis, they should be included on the grant and/or the resulting journal article.

The following references are excellent sources for additional information about the statistical topics described in this article.

Literature Cited and Selected References

- Bolker, B.M., M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M.H. Stevens, and J.S. White. 2009. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends Ecol. Evol.* 24:127–135.
- Cochran, W.G. and G.M. Cox. 1957. *Experimental designs*. 2nd ed. Wiley, New York, NY.
- Cohen, J. 1992. A power primer. *Psychol. Bull.* 112:155–159.
- Gelman, A. and E. Loken. 2014. The statistical crisis in science: Data-dependent analysis—a “garden of forking paths”—explains why many statistically significant comparisons don’t hold up. *Amer. Sci.* 102:460.
- James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An introduction to statistical learning*. Springer, New York, NY.
- Keselman, H.J. 2015. Per family or familywise Type I error control: “Eether, eyether, neether, nyther, let’s call the whole thing off!” *J. Mod. Appl. Stat. Methods* 14(1):6.
- Kramer, M.H., E.T. Paparozzi, and W.W. Stroup. 2016. Statistics in a horticultural journal: Problems and solutions. *J. Amer. Hort. Sci.* 141:400–406.
- Lang, T.A. and M. Secic. 2006. *How to report statistics in medicine: Annotated guidelines for authors, editors and reviewers*. 2nd ed. American College of Physicians, Sheridan Press, Chelsea, MI.
- Little, T.M. 1978. If Galileo published in *HortScience*. *HortScience* 13:504–506.
- Milliken, G.A. and D.E. Johnson. 2009. *Analysis of messy data*. Vol. 1, 2nd ed. Chapman & Hall/CRC Press, Boca Raton, FL.
- Nakagawa, S. and I.C. Cuthill. 2007. Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biol. Rev. Camb. Philos. Soc.* 82:591–605.
- Onofri, A., E.A. Carbonell, H.-P. Piepho, A.M. Mortimer, and R.D. Cousens. 2010. Current statistical issues in *Weed Research*. *Weed Res.* 50:524.
- Reinhart, A. 2015. *Statistics done wrong: The woefully complete guide*. No Starch Press, San Francisco, CA.
- Schabenberger, O. and F.J. Pierce. 2002. *Contemporary statistical models for the plant and soil sciences*. CRC Press, Boca Raton, FL.
- Stroup, W.W. 2013. *Generalized linear mixed models: Modern concepts, methods and applications*. CRC Press, Boca Raton, FL.
- Stroup, W.W. 2015. Rethinking the analysis of non-normal data in plant and soil science. *Agron. J.* 107:811–827.
- Tressoldi, P.E., D. Giofré, F. Sella, and G. Cumming. 2013. High impact = high statistical standards? Not necessarily so. *PLoS One* 8(2):E56180, doi: 10.1371/journal.pone.0056180.
- Vance, E.A. 2015. Recent developments and their implications for the future of academic statistical consulting centers. *Amer. Stat.* 69:127–137.
- Wasserstein, R.L. and N.A. Lazar. 2016. The ASA’s statement on *p*-values: Context, process, and purpose. *Amer. Stat.* 70:129–133.
- Weissgerber, T.L., N.M. Milic, S.J. Winham, and V.D. Garovic. 2015. Beyond bar and line graphs: Time for a new data presentation paradigm. *PLoS Biol.* 13(4):E1002128, doi:10.1371/journal.pbio.1002128.