Statistics in a Horticultural Journal: Problems and Solutions

Matthew H. Kramer[1]
U.S. Department of Agriculture, Agricultural Research Service, Statistics Group, Building 005, Room 130, 10300 Baltimore Avenue, Beltsville, MD 20705

Ellen T. Paparozzi
Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE 68583-0724

Walter W. Stroup
Department of Statistics, University of Nebraska-Lincoln, Lincoln, NE 68583-0963

ABSTRACT. We examined all articles in volume 139 and the first issue of volume 140 of the *Journal of the American Society for Horticultural Science* (*JASHS*) for statistical problems. Slightly fewer than half appeared to have problems. This is consistent with what has been found for other biological journals. Problems ranged from inappropriate analyses and statistical procedures to insufficient (or complete lack of) information on how the analyses were performed. A common problem arose from taking many measurements from the same plant, which leads to correlated test results, ignored when declaring significance at $P = 0.05$ for each test. In this case, experiment-wise error control is lacking. We believe that many of these problems could and should have been caught in the writing or review process; i.e., identifying them did not require an extensive statistics background. This suggests that authors and reviewers have not absorbed nor kept current with many of the statistical basics needed for understanding their own data, for conducting proper statistical analyses, and for communicating their results. For a variety of reasons, graduate training in statistics for horticulture majors appears inadequate; we suggest that researchers in this field actively seek out opportunities to improve and update their statistical knowledge throughout their careers and engage a statistician as a collaborator early when unfamiliar methods are needed to design or analyze a research study. In addition, the ASHS, which publishes three journals, should assist authors, reviewers, and editors by recognizing and supporting the need for continuing education in quantitative literacy.

The incorrect use of statistics in scientific articles seems to be a never-ending discussion topic. A current controversy involves a decision by *Basic and Applied Social Psychology* in 2015 to ban the use of *P*-values (i.e., null hypothesis testing) in articles appearing in their journal. This prompted the American Statistical Association to publish, in 2016, a policy statement on the use of *P*-values in research publications. Reinhart (2015) in his book, *Statistics Done Wrong: The Woefully Complete Guide*, gives a good overview of the sorts of statistical mistakes made in science, with many biological examples.

There are also attempts to gauge how severe the misuse of statistics is in various biological disciplines. The article on the website hosted by influentialpoints.com (Dransfield and Brightwell, 2012) provides an overall guide to statistics misuse in biology, with a bias toward medicine. The authors of this site categorized errors found in an examination of "several thousand papers" and the article posted is abstracted from their book (Brightwell and Dransfield, 2013).

A recent evaluation of incorrect analyses of interaction effects in the neurosciences found that about half the published articles had statistical issues when analyzing factorial treatment designs, with some apparently severe enough to call the study's conclusions into question (Nieuwenhuis et al., 2011). A recent *Nature* article by Allison et al. (2016) discussed how easy it was to find mistakes in data handling in publications, but how hard it was to get them fixed. Although there are many reasons why a statistical analysis may or may not be appropriate, only those most applicable to horticulture will be discussed below.

We examined issues of the *JASHS* published between Jan. 2014 and Jan. 2015 inclusive, for statistical problems. This was prompted by an interest in revising the currently antiquated instructions to authors about the use of statistics in the society's journals. To do this, we needed to identify the kinds of statistical methodologies required by current authors to support their findings, the kinds of data being collected, and what authors were actually doing when analyzing the data. The revised version of statistics instructions will be appearing separately. Here, we describe the kinds of statistical errors most commonly made by authors in this journal and characterize the patterns of errors and omissions we found. These are not necessarily fatal flaws, but reveal weaknesses that may affect conclusions. We then ascribe probable causes and suggest some possible remedies. We hope this review will be helpful both to authors and reviewers.

Methods

Eighty-six articles from *JASHS* (all issues in 2014 plus issue 1 in 2015) were examined to characterize the kinds of statistical methodology used and associated problems. This involved reading each article to understand the primary objectives of the research, decide if appropriate statistical methodology was applied, and identify any statistical issues associated with the handling of the data. In some cases, insufficient information was provided to

400

J. Amer. Soc. Hort. Sci. 141(5):400–406. 2016.

understand a study's data analysis. This complicated our job when trying to determine if the data were correctly analyzed and thus, decide if there were problems. In other cases, there was no mention of statistical methods, yet the Results section clearly indicated that the data were analyzed statistically in some way, so clues were sought in the text, figures, or tables. Failure to describe the statistical methodology used is of itself a serious statistical issue, and something that should not occur in a refereed journal article. This was tabulated as such. Many journals now require authors to archive their raw data and computer code, some in public domain databases, others with the journal. Examples of journals that require at least some data archiving include *The American Naturalist*, *BMC Ecology*, *Genetics*, *Molecular Ecology and Evolution*, *Nature*, and *Science* (UC3 Data Pub Blog, 2012). It is not inconceivable that in the near future some fact checking by reviewers will be accomplished by verifying that the statistical code used for the analysis on the raw data are both appropriate and produces the stated results.

After each article was read, if a statistical issue was found, it was briefly summarized. These summaries were tabulated and used to develop a categorization scheme to identify key issues. The statistical software used for each article was also noted, as a way to understand how current horticulture researchers use statistical software.

Results and Discussion

This section is divided into three subsections. In the first subsection, we describe the statistical problems that were found, and briefly explain why they matter. In the second subsection, we list the statistical computing software used to implement analyses, problems associated with reporting, and software choices. In the third subsection, we postulate various reasons for why these problems arose.

STATISTICAL PROBLEMS. Table 1 shows a summary of the statistical problems that were found. The most common problem (30 articles) was inappropriate analysis of data from multiple dependent variables on the same unit of observation. Specifically, variables were analyzed one at a time with no attempt to account for between-variable correlation and no attempt to control for experiment-wise error rate; i.e., the likelihood of making at least one type I error when two or more tests are performed. The latter is similar to the issue of multiple comparisons of treatments, where results from hypothesis tests are correlated (Westfall and Young, 1993), discussed in more detail below. In other words, if more than one kind of measurement is made on each plant (say fruit yield and mean fruit sugar content), then the two measures cannot both be independently tested at $\alpha = 0.05$. Measurements are independent only if they are made on *different* plants. Obviously, requiring a different plant for each response variable would be both impractical and prohibitively expensive. The reality is that multiple response variables are often measured on the same plant. This is a valid design approach, but it does require an analysis that accounts for correlation among measurements. For example, if one of the plants is nitrogen deficient, it is likely that both its fruit yield and its fruit sugar content would be affected. Failure to account for this kind of correlation can distort findings in a number of ways. A treatment effect may exist, and be detectable when correlated variables are analyzed together using a multivariate analysis, whereas one-at-a-time testing can mask the effect. On the other hand, separate analyses for each response variable can make the tests too liberal, because one is assuming the tests are independent when they are not. See Hochberg and Tamhane (1987) for a discussion of multivariate issues and Johnson and Wichern (2007) for a complete presentation of multivariate analysis. Horticultural researchers need to be aware of this issue and learn how to deal with it. As a final point, often the correlations between the dependent variables are of intrinsic interest, as groups of variables may respond similarly when faced with environmental changes or if different cultivars are used. In fact, building networks of fruit characteristics or plant metabolites is based on this assumption (Fatima et al., 2016). When correlation is disregarded in statistical analysis, important information about relationships among the dependent variables is lost.

The next category of problems (24 articles) had some other kind of incorrect analysis (itemized in Table 2) other than means separation problems, which we discuss separately below (Table 3). These problems had to be obvious for us to identify them, since the raw data were not available. The two most common types of problems characterized in Table 2 were as follows. In 11 cases, inspection of the figures revealed an obvious relationship between the mean and variance. Typically, the variance increased with larger means, yet the statistical analysis used a method that requires the assumption of no mean-variance relationship. This suggests a larger problem of failure to verify assumptions. Given that we have no way of knowing whether the statistical assumptions underlying most of the tests reported were satisfied, it is likely we actually have an undercount of the true number of articles with these types of problems. In our consulting experience with biological researchers (M.H. Kramer and W.W. Stroup), we find that many researchers are not aware of the underlying assumptions, how to test for them, or how to perform postanalysis model diagnostics. The second most frequent problem listed in Table 2, (seven instances) concerned inconsistencies between how the data were described (the study design)

Table 1. Summary of identified statistical problems found in 86 articles published in the *Journal of the American Society for Horticulture Science*. One article may have more than one problem identified.

Problem	Count
Need experiment-wise control/multiple dependent variables	30
Incorrect analysis	24
Means separation	20
Missing information	10
Miscellaneous	8

Table 2. Specific incorrect analysis methods found in 24 of 86 articles published in the *Journal of the American Society for Horticulture Science*.

Problem	Count
Variance a function of mean	11
Random effect treated as fixed or ignored	7
Ignored spatial variability	1
Repeated measures ignored	1
Wrong repeated measures covariance structure	1
Pooled different treatments	1
Ignored censoring	1
Regression with three observations	1

and how they were analyzed. For example, there may have been constraints on the randomization of the observations, such as blocking in a randomized complete block design, by locations (plots of land) or by occasions (different years), but the analysis used a method that failed to account for these sources of variation.

Incorrect means separation procedures (20 articles) occurred in a variety of forms (Table 3). Different means separation procedures can produce different groupings of means (Day and Quinn, 1989). Some means separation procedures (e.g., the Scheffé and Bonferroni tests) are specifically intended to be used when the consequences of type I error (falsely concluding a treatment effect exists) are considered especially serious, whereas other tests (e.g., the Duncan or Tukey) are specifically intended to be used when the consequences of a type II error (failing to detect a non-negligible treatment effect) are considered more serious. Control of error rates is very important in genomic studies, where there may be millions of comparisons, all using the same few individual organisms (here error rates are often controlled using the false discovery rate method, see Benjamini and Hochberg, 1995). Error control is a complex issue, because controlling type I error increases the chance of making a type II error, and vice versa. Achieving the right balance between the two at the design stage requires some thought. However, we found no indication that any effort went into finding this balance. See Chapter 3 in Milliken and Johnson (2009) for a complete discussion and recommendations concerning multiple comparison procedures.

Because the choice of method could affect conclusions about treatments, researchers must be explicit about what mean separation method was used and the rationale for using it. An equally important point is that mean separations tests only identify *which* treatments are different. They do not provide sufficient information about *how* different. This requires a confidence interval, or at least a properly estimated standard error of the *difference* (not a standard error of the mean—they are not interchangeable). The standard error of the mean allows one to determine a confidence interval for the mean—period. The standard error of the difference is the quantity used when testing if treatment means differ or obtaining a confidence interval for the treatment difference, often the objective of an experiment. In many common designs (e.g., any design with blocking), there is no straightforward way to determine the standard error of the difference from the standard error of the mean. Providing only the standard error of the mean is a form of misrepresenting the data, because if readers try to use the standard error of the mean to calculate a standard error of the difference—and they will—and there is blocking, they will get it wrong, opening the prospect of readers misinterpreting research results. Relevant information about the treatment difference is usually the most important information available from the research data, and unfortunately rarely provided. See Littell et al. (2006) for a complete discussion of the standard error issue.

In five of the 20 articles, the method of means separation was not given. Other problems included no adjustment for multiple comparisons and no accompanying rationale for not doing so and mean separation that was apparently performed without a prior analysis of variance (ANOVA).

In the next category, "missing information," with 10 articles, the explanation of how the analysis was done was either absent or so vague that we could not figure out what methods were used, even after looking through the figures and Results section (Table 4). Clearly, these analyses could not be reproduced. Indeed, one generally needed improvement in articles is to provide sufficient information about how the data are collected and handled so that others could reproduce the analysis if given the same raw data. This should be considered a failure of the review process and should not occur in a refereed journal article.

The remaining category, "miscellaneous," with eight articles, had other problems that did not fit into one of the above categories (Table 5), such as not reporting sample size, or an inconsistency between what we knew the software to do and how the authors reported using it.

SOFTWARE PACKAGES. Out of 86 articles, 10 used no statistics, 57 used one package/program, 10 used two, and nine used three or more. Overall, there were 39 different programs used (seven articles did not name the software used). Ten were "general use" programs [e.g., SAS (SAS Institute, Cary NC), JMP (SAS Institute), R (R Core Team, 2013)], used in 62 articles, and the rest "specialty" programs (largely for genomics or phylogenetics), used in 42 articles. Details are provided in Tables 6 and 7. SAS was by far the most widely used general statistics package. Authors and reviewers should recognize that statistical software is a means of implementing a statistical analysis, not a statistical *method* in itself. Problems occurred when the statistical method was given, but not the software used to implement it or vice versa. Sometimes a method was given, but the software used was clearly not capable of implementing the analysis described (e.g., use of SAS PROC GLM to analyze data with random model effects). Note that although PROC

Table 3. Problems with means separation procedures found in 20 of 86 articles published in the *Journal of the American Society for Horticulture Science*.

Problem	Count
Duncan's used for means separation	8
Undisclosed means separation technique	5
No adjustment for multiple comparisons (e.g., used *t* tests)	4
Means comparisons without prior ANOVA	2
Used nonoverlapping confidence intervals as means comparison	1

ANOVA = analysis of variance.

Table 4. Problems due to missing information in 10 of 86 articles published in the *Journal of the American Society for Horticulture Science*.

Problem	Count
Missing necessary statistical information	7
Not clear what statistical software was used for	1
Undisclosed tests	1
PCA[z] results not explained adequately	1

[z]Principle component analysis.

Table 5. Miscellaneous statistical problems found in 8 of 86 articles published in the *Journal of the American Society for Horticulture Science*.

Problem	Count
Sample sizes not given	3
Measure of variability not reported	2
SAS[z] PROC MIXED does not do stepwise variable selection	1
Show just fitted curves	1
Figure issues	1

[z]SAS Institute, Cary, NC.

GLM does have a random statement, limitations in its ability to obtain correct statistics for tests and confidence intervals were the primary motivation for developing PROC MIXED and GLIMMIX. For example, with PROC GLM, means separation uses estimates from an all fixed effects model regardless of whether the random statement is used or not.

Many of the problems we have identified are in areas where statistical software development is in its infancy. One example involves multiple measures on the same plant that are correlated, but some are qualitative and some are quantitative. However, improved methodology and associated software are likely to become available in the future, hence the need for continuing education in statistics.

UNDERLYING REASONS FOR THESE PROBLEMS. Years ago, Gates (1991) and Little (1978) documented some of the same problems reported above, including problems with means separation methods similar to those we describe, and focusing on the disconnect between how experiments were conducted and how they were analyzed. These problems are not unique to horticulture. We know from discussions with our colleagues at national meetings dedicated to statistics in agriculture that many of the problems we found exist in other biological disciplines. Why do these problems occur? Why do they persist? Have efforts over the past 25 years to address these issues been ineffective? Do we need to rethink our approach to statistical practice and reporting? In this section, we suggest reasons for the statistical issues discussed above. The next section presents recommendations.

We begin by considering what is currently available. There are ample written materials that provide statistical methodology guidance for biologists. For example, an Amazon.com (Seattle, WA) search on "statistics biology," done 11 Mar. 2015, brings up 3785 results. Many of these are books with material on common issues in horticultural research. Although emphases differ, many of these books are written explicitly with biological researchers as the target audience. Statistical methods courses are an integral part of the training that most horticultural researchers receive. Both land-grant universities and U.S. Department of Agriculture's (USDA) Agricultural Research Service (ARS) have some form of statistical consulting capability. ASHS has statistical editors, who act as a resource if an editor or other reviewers flag an article as needing statistical review. We do not believe that scientists in horticulture are less statistically savvy than researchers in the other biological sciences. Yet these problems occurred in a high proportion of articles that we examined.

In considering possible reasons why these problems occur, we suggest five main themes: 1) rapid changes in both horticultural and statistical science; 2) demands on time vs.

Table 6. Categories and counts of the particular statistical software packages used in 86 articles published in the *Journal of the American Society for Horticulture Science*. One article may identify more than one program.

Category	Unique software programs	Count (articles with)
General use	10	62
Specialty	29	42
Unknown	?	7
None	—	10

? = From the description the author presented, it was hard to tell if the software was unique.

Table 7. Frequency of general and specialty statistics programs used in 86 articles published in the *Journal of the American Society for Horticulture Science*. One article may identify more than one program. All software packages can be found by conducting a web search for the identified program.

General	Count	Company	Specialty[z]	Count	Company
SAS	35	SAS Institute Cary, NC	MEGA software	4	Biodesign Institute, Tempe, AZ
JMP	8	SAS Institute	NTSYS-PC	4	Exeter Software, Setauket, NY
SPSS	5	IBM Armonk, NY	Structure software	4	Pritchard Laboratory, Stanford University, Stanford, CA
R	4	R Core Team <https://www.r-project.org/>	ASReml	3	VSNi Hemel Hempstead, UK
GenStat	3	VSNi, Hemel Hempstead, UK	GeNorm software	2	Schlotter et al. (2009) <https://genorm.cmgg.be/>
CoStat	2	CoHort Software <http://www.cohort.com/>	NormFinder	2	Molecular Diagnostic Laboratory, Aarhus University Hospital, Aarhus N, Denmark
Statistica	2	Dell, Round Rock, TX			
Minitab	1	Minitab, State College, PA			
Statgraphics	1	StatPoint Technologies, Warrenton, VA			
InfoStat	1	National University of Cordoba, Cordoba, Argentina			

[z]Other specialty programs used once: JoinMap and MapQTL (Kyazma, Wageningen, the Netherlands); LOD model (Valve Developer Community <https://developer.valvesoftware.com/wiki/LOD_Models>); BestKeeper software (<http://www.gene-quantification.com/bestkeeper.html>); DNA Manipulation software (Bioinformatics Organization, Alberta, Canada); KinGroup (FoxToo <http://www.foxtoo.com/Mac/download-KINGROUP-10059935.htm>); Populations software (<http://bioinformatics.org/~tryphon/populations/>); Tree Explorer bio software (BioSoft Net <http://en.bio-soft.net/page.html>); TableCurve3-D (SysStat Software, San Jose, CA); QTLnetwork (<http://ibi.zju.edu.cn/software/qtlnetwork/>); Cerus software PowerMarker (<http://statgen.ncsu.edu/powermarker/index.html>); GenAIEx (<http://biology-assets.anu.edu.au/GenAlEx/Welcome.html>); Splitstree software (<http://www.splitstree.org/>), GDA software (<http://en.bio-soft.net/dna/gda.html>); GAPIT-R (Buckler Laboratory for Maize Genetics and Diversity, Ithaca, NY); DARwin software (<http://darwin.cirad.fr/>); qBase software (qBase,Washington, DC); FreeTree software (<http://web.natur.cuni.cz/flegr/programs/freetree.htm>); SimaPro software (<http://www.simapro.co.uk/>); Custom Statistics (<https://confluence.jetbrains.com/display/TCD9/Custom+Statistics>); Excel add-ons (Microsoft, Seattle, WA).

the need to stay current; 3) the current state of statistical education; 4) the review infrastructure; and 5) the current model for horticulturist–statistician interaction. These will be discussed in the order in which they are listed.

Horticultural and statistical sciences are both changing rapidly. In particular, statistics is not a static set of algorithms; it evolves over time just like any other area of science. Methodology accepted 20 years ago may be considered antiquated or unacceptable now. One good example of this is the concept and implementation of random effects, such as blocks and studies at multiple locations or occasions. These factors were typically modeled as fixed effects until software became available in the early 1990s to model them correctly as random effects. Another example is the use of transformations of dependent variables when the assumptions of ANOVA or regression were violated. Using generalized linear mixed models (GLMMs), especially for dependent variables with non-normal distributions, is demonstrably more accurate than transformations. However, usable GLMM software has only appeared in the past decade. A third example is the increasing use of the Bayesian framework for modeling data.

A major factor driving changes in statistical practice is statistical software. With just a few mouse clicks, one can compute all kinds of statistics and tests that appear to be bona fide, even if the model is conceptually inappropriate. Software is not going to spontaneously protest about what it has been asked to do, or tell the user, "perhaps you should consider a more suitable alternative." Contemporary software and methodology advances such as GLMMs and Bayesian approaches offer more accurate and insightful analysis, but also require computational resources that were unthinkable as recently as a decade ago. They also have a greater potential for abuse if not used with the requisite understanding. These are but a few examples of the changes occurring in statistics for the biological sciences.

Researchers and reviewers are thus caught in a bind. On one hand, they need to keep current with these advances in statistical practice. Older methods still produce "statistics" and P-values, but the utility of these analyses is increasingly compromised relative to newer, better alternatives. For this reason, we advise seeking statistical advice only from researchers who actively follow changes in statistical practice. We will expand on this point later in this section. On the other hand, while it is easy to say, "Researchers need to keep current," it is quite another thing to actually do so. Keeping up is a challenge, especially when one's primary discipline is horticulture, not statistics. One very real problem that we all face as professionals is competing demands for time. It can be tempting to simply take a program used for a previous study and rerun it, substituting a new data set for an old one. This is not usually a recipe for success.

A third factor contributing to these problems is that introductory statistical methods courses may not teach what students need for their careers. From the viewpoint of biological science students, these classes present a lot of unfamiliar material in a short amount of time. Many students are uncomfortable with the mathematics essential to statistical methodology. Students have to learn a statistical programming language, often their first exposure to writing computer code. At the same time they take introductory statistics, students are learning the literature of their field, and as a result, may not appreciate the importance of statistics, much less the kinds of statistics they are likely to use and why. Consequently, many students report difficulty seeing how information presented in their introductory statistics class

will be relevant to their research. From the instructor viewpoint, making curriculum decisions for such classes is not easy. Instructors must take into account the very heterogeneous backgrounds and needs of the students who take these classes. Because this may be the only formal statistics training these students have for their entire career, instructors need to condense an entire field of study into one or two semesters. In many cases, topics that students are likely to need—e.g., analyzing multiple measurements on the same plant—are not covered because the material is considered too complex given limitations in student proficiency and confidence in math and computing. All of this makes it unrealistic to expect that, upon completion of their statistics classes, students will have the ability to correctly design and conduct an experiment, analyze the data, and interpret the results. It would be more realistic to expect that they leave with the ability to converse effectively with a collaborating statistician and to have sufficient background to do some investigation of methods they were not exposed to on their own. Even the best students cannot learn enough in two semesters to be prepared for their career. Students as professionals must become life-long learners.

Closely related to method courses are method textbooks. There are a great many statistics texts aimed at biologists, and they do not necessarily share a common core of concepts. Some concepts important for horticultural researchers, such as correlated variables measured on the same plant, are rarely included. A horticulturist who is under a time crunch and trying to determine what to do may be overwhelmed by the information presented in a good textbook and struggle to identify methods appropriate to analyze the data at hand.

A fourth factor is the review process. Reviewers or referees of articles are usually chosen for their subject matter knowledge; they are peers in that scientific field and not statisticians. As a result, they may not be current on good statistical practice, especially if the authors are using a recently developed or infrequently used method. Some reviewers may accept on faith that the appropriate statistical method has been used, the modeling is correct, assumptions of the model satisfied, and fail to catch statistical errors. In addition, statistical review is severely constrained when authors do not give sufficient statistical details, data are not presented in figures, and results are limited to variables whose tests were "significant." In other cases, reviewers whose knowledge on statistics is dated may provide well-meaning, but inappropriate feedback regarding statistical aspects of the manuscript being reviewed, or, even worse, incorrectly reject the manuscript thinking the statistics are flawed when in fact they are legitimate, just not understood by the reviewer.

Although it would be desirable to have a statistician review the statistical aspects of journal submissions, the reality is that there are not enough statisticians in the world to review every manuscript submitted to biological journals, nor would most statisticians be interested in spending their time providing such a service, even if rewarded for doing so. Thus, for the most part, biologists themselves must provide this service, which requires statistical expertise, both when conducting experiments and when reviewing journal articles. There must be a balance between what a researcher should know about statistics and knowing when it is time to consult with a statistician. This balance depends on an individual researcher's knowledge of statistics.

Finally, there is the most important, and perhaps the most difficult issue, the way in which we approach the interaction between horticultural researcher and statistician. There are two

404

J. Amer. Soc. Hort. Sci. 141(5):400–406. 2016.

predominant models for this interaction. One is the "home repair" model: try doing it yourself until/unless you get in over your head, then see a statistical consultant. The other is the "dry cleaner" model: drop your data off at the statistical consulting center, explain what you need, and pick up the results, possibly including a write-up, when they are ready. Notice that both models conceptualize the role of the statistician as consultant or technician. There are two problems with both models. First, they only engage the statistician with the technical aspects of data analysis, not with the scientific question that provides the context for the study, the way it was designed, the data it produced, and the larger goals of the analysis. The "dry cleaner" model compounds the situation by disengaging the horticulturist from analysis of the data. Detaching statistics from the science increases the likelihood of the kinds of problems we found in the articles we examined. Second, while tenure-track statistics faculty at some land-grant universities once received credit in merit evaluations for consulting, as long as it led to publication, this is no longer the case. Collaboration, yes, consulting, no. We continue this discussion in the "Recommendations" section.

What are the consequences of incorrect analysis? Would the study's conclusions change? In some cases they would not. If a scientist plots the raw data and the effect of a treatment is large, even the wrong analysis will likely bring one to the right conclusion (i.e., the results are obvious, even if one used no statistics). However, we are long past the time of large effect sizes being typical (think back to early experiments demonstrating that fertilization improved yield); as science matures there tends to be more whittling away at the edges and less carving. In experiments with smaller effect sizes, the wrong analysis will more likely lead one astray, perhaps concluding that treatments differ when they do not, or vice versa. This can have a subsequent biological or economic cost, for example, selecting a genotype that later fails to perform as predicted.

Recommendations

We suggest three areas of focus for horticulturists to improve the accuracy of their use of statistical science: continuing education, collaboration, and communication.

CONTINUING EDUCATION. Given that it is not possible for a horticulturist to learn, during graduate school, all the statistics—or horticulture—needed over an entire career, life-long learning is essential. Continuing education should become a part of the researcher's diet. This could take many forms. We suggest five practical, easily implemented beginning steps: annual statistical updates at national meetings, auditing statistical design and analysis classes at their respective institutions, inviting statisticians at their respective institutions to give seminars or tutorials about statistical methods for horticulture, and/or visiting statistical websites such as JMP at SAS Institute to view video updates. Also, many universities are also starting to produce video tutorials, often short and focused on specific statistical issues. ASHS should work with North Central Coordinating Committee (NCCC)-170, a USDA-sponsored consortium of statisticians from land-grant universities and ARS, to make these resources known and available to horticultural scientists.

COLLABORATION. In a previous draft of this paper, we recommended planning: plan before you plant! This is crucial: the statistical thinking that goes into planning a study—before any data are collected—whether it is a formally designed experiment, a survey, or an observational study, is the most important use of statistics in research. Think about your objectives before you visit with your collaborating statistician. Your results will only be as good as the design and analysis. To be fully effective, this recommendation goes beyond planning. Few, if any, of the statistical problems we found would have occurred if a statistician had been engaged as a collaborator in research. Consulting is an isolated act to solve a specific statistical problem; collaboration is a partnership to address a scientific question. Bringing your data to the statistician after the experiment has been conducted is consulting in its least effective form, and is an open invitation to problems.

In discussing statistical education, we concluded by saying that it is unrealistic to expect a horticultural graduate student to learn all the statistics needed over an entire career, and that students must be prepared to become life-long learners. Although we strongly encourage continuing education, we do so with the caveat that the "home repair" model described earlier, while sometimes necessary, is often insufficient. Contemporary scientific research is too complex and multidisciplinary to be done without involving expertise from all relevant disciplines. This raises the problem of human resources. In a perfect world, a statistician should be a fully engaged collaborator from the inception of every research project. While this is the ideal, it is also impractical. As a result, we strongly urge horticulture and statistics to look to the future and encourage and support partnering doctoral students. This would not only improve the quality of statistics in horticulture, but also it would improve science literacy among statisticians. More importantly, it would teach scientists-in-training the art of collaboration at a time when they are most likely to derive career-long benefits from the experience.

COMMUNICATION. As a first step, horticulturists should review the soon to be published ASHS Statistical Guidelines for Authors. In a research manuscript, authors, at a minimum, need to include enough information on their experiment design (e.g., levels of blocking or other constraints on randomization that induce correlation, units to which treatment are assigned), treatment design (e.g., treatment factors and their levels or categories), and method of analysis to allow for a fair review. Given this information and the raw data, a reviewer should be able to reproduce all important aspects of the analysis. Any paper with quantitative data that does not address these three items should be returned to the author. Chapter 5 of Milliken and Johnson (2009) defines and explains experiment and treatment designs.

In order for ASHS to communicate the importance of statistics in articles published in their journals, we suggest that the society consider some form of informal credentials in order for a reviewer/editor to comment on the statistics. For example, a reviewer could be encouraged to take a statistics short course/workshop/tutorial (ideally one created for reviewers and editors) every 5 years. To accomplish this, ASHS must also commit to providing continuing education to the membership via venues such as regional or national meetings, webinars, or other opportunities. For many universities, this could also fulfill a faculty member's professional development requirement.

Summary and Conclusions

We found statistical issues in about half of the articles published in *JASHS* from Jan. 2014 to Jan. 2015. This finding is

J. Amer. Soc. Hort. Sci. 141(5):400–406. 2016.

405

not unique to this time period, nor is it new. Discouragingly, Gates (1991) reported similar levels of problems for publications in horticulture. Problems exist and they persist. This suggests that the current way of doing things with regard to statistics in horticulture is not working and needs rethinking. Both disciplines, horticulture and statistics, have a role to play.

On the statistics side, we know that efforts are underway to rethink the content and approach of methods courses taught to biological science graduate students. We support and encourage these efforts. We also support and encourage efforts to make researcher-friendly continuing education materials available, especially those that can be accessed online.

On the horticulture side, we make three recommendations: continuing education, collaboration, and communication. Horticulture and statistics are both changing rapidly, and will continue to change. Life-long learning is essential and this means a commitment to workshops and tutorials in statistics specifically tailored to the needs of horticultural researchers, and a commitment by horticulturists to take advantage of these opportunities. Modern research is inherently multidisciplinary. The persistent number of errors in horticulture publications is evidence—unwelcome, perhaps, but real—that the statistician-as-occasional-consultant model is not working and needs to be replaced by genuine collaboration. Given the reality of the number of available statisticians and the demands on their time, collaboration necessarily requires involving doctoral graduate students. Finally, the review process needs attention. Our survey of recent journal publications serves as stark evidence that errors are getting through peer review. In addition, we know that well-meaning but misguided suggestions based on inadequate or dated statistical knowledge also occur. We suggest continuing education specifically focused on reviewer and editor needs to address these issues.

Literature Cited

Allison, D.B., A.W. Brown, B.J. George, and K.A. Kaiser. 2016. Reproducibility: A tragedy of errors. Nature 530:27–29.

Benjamini, Y. and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. B. 57:289–300.

Brightwell, R. and R.D. Dransfield. 2013. Avoiding and detecting statistical malpractice: Design and analysis for biologists, with R. 25 May 2016. <http://influentialpoints.com/aboutus.htm>.

Day, R.W. and G.P. Quinn. 1989. Comparisons of treatments after an analysis of variance in ecology. Ecol. Monogr. 59:433–463.

Dransfield, R.D. and R. Brightwell. 2012. Statistical mistakes in research: Use and misuse of statistics in biology. 16 Nov. 2015. <http://influentialpoints.com/Training/statistical_mistakes_in_research_use_and_misuse_of_statistics_in_biology.htm>.

Fatima, T., A.P. Sobolev, J.R. Teasdale, M. Kramer, J. Bunce, A.K. Handa, and A.K. Mattoo. 2016. Fruit metabolite networks in engineered and non-engineered tomato genotypes real fluidity in a hormone and agroecosystem specific manner. Metabolomics 12:103.

Gates, C.E. 1991. A user's guide to misanalyzing planned experiments. HortScience 26:1262–1265.

Hochberg, Y. and A.C. Tamhane. 1987. Multiple comparison procedures. Wiley, New York, NY.

Johnson, R.A. and D.W. Wichern. 2007. Applied multivariate statistical analysis. 3rd ed. Pearson, New York, NY.

Littell, R.C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberger. 2006. SAS for mixed models. 2nd ed. SAS Institute, Cary, NC.

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: Designed experiments. Vol. 1, 2nd ed. Chapman and Hall/CRC, New York, NY.

Nieuwenhuis, S., B.U. Forstmann, and E.J. Wagenmakers. 2011. Erroneous analyses of interactions in neuroscience: A problem of significance. Nat. Neurosci. 14:1105–1107.

R Core Team. 2013. R: A language and environment for statistical computing. 25 May 2016. <http://www.R-project.org/>.

Reinhart, A. 2015. Statistics done wrong: A woefully complete guide. No Starch Press, San Francisco, CA.

Schlotter, Y.M., E.Z. Veenhof, B. Brinkhof, V.P. Rutten, B. Spee, T. Willemse, and L.C. Penning. 2009. A GeNorm algorithm-based selection of reference genes for quantitative real-time PCR in skin biopsies of healthy dogs and dogs with atopic dermatitis. Vet. Immunol. Immunopathol. 129:115–118.

UC3 Data Pub Blog. 2012. Archiving data, best practices, data sharing. 20 Apr. 2016. <https://datapub.cdlib.org/2012/11/20/thanks-in-advance-for-sharing-your-data/>.

Westfall, P.H. and S.S. Young. 1993. Resampling-based multiple testing: Examples and methods for P-value adjustment. Wiley, New York, NY.

406

J. Amer. Soc. Hort. Sci. 141(5):400–406. 2016.