

SILVER STATE-ISTICS



Nevada Chapter of American Statistical Association

Nevada Chapter Newsletter

Vol. 14 No. 1

Dear Nevada ASA Members and Friends.

Read on for information on what our Chapter has been up to and what we are planning for the future. Please keep an eye on our website for more details on all upcoming events.

www.amstat.org/chapters/

Contents:

Coming Events	1
From the President's Desk	1
Fall Symposium 2016	2
New Leaders of NV-ASA	3
Statistics Journals and Books Looking for a	
Home	3
Nevada State College News	3
UNLV News	4
UNR News	4
More on P-Values	4
Joining NV-ASA!	5
NV-Chapter Officers & Others	5

* * *

Coming Events

Our annual K-12 Poster Competition is underway. Submissions are due in late March; results should be announced in mid-April.

Our **Annual Meeting** will take place in the North this coming Fall; stay tuned for announcements of this and other events.

Elections: Positions that will be up for elections this coming fall are President, Northern Vice President, and Secretary.

From the President's Desk

Debra Stiver

Best wishes for a happy and productive New Year!

Nevada ASA begins 2017 by expressing our thanks to our retiring officers and welcoming the incoming officers. Chad Cross has served as the Nevada ASA Chapter Representative to national ASA for the past 3 years and has been a long-time contributing member in chapter activities. Also retiring are Guogen Shan as Treasurer from Southern Nevada and Erick Gremlich as the Northern Nevada Vice President. The chapter could not function without member participation in these important roles. It is with gratitude that we bid them adieu from their official positions. We will see them all in the chapter activities currently being planned for 2017.

The retiring officers handed over their duties to the new slate of officers at the NV-ASA Annual Meeting and Symposium, held at UNLV this past October. We welcome Anna Panorska as Chapter Representative, Chris Tong as Treasurer, Alejandra Livingston as Northern Nevada Vice President, and Charles Davis, entering his second term as the Southern Nevada Vice President. The new board consists of members who work in a wide variety of statistical professions including academia, the federal government, state government, and private consulting. It mirrors the wide variety of interests represented by the American Statistical Association.

I would like to give a special thanks to Dave Thiel, founder and coordinator of the NV-ASA K-12 Poster Competition. This successful outreach activity has benefitted students from all age groups for many years. Introducing students to the statistical sciences is one of the primary missions of ASA, and Nevada has been fortunate to have Dave at the helm. He is moving on to new adventures, and Tia Price from Southern Nevada has taken over the duties of coordinating the contest. She has been a competition judge for many sessions. NV-ASA looks forward to continuing the competition with Tia stepping in as coordinator.

We invite you to participate in our chapter and national activities regardless of your statistical interest – from aspiring student to accomplished professional! 2017 board meetings began in January. Look for announcements of this year's upcoming technical and social events in emails and future newsletters. •

Charles Davis

Our Fall Symposium with Annual Meeting took place Saturday, October 22 on the UNLV campus, co-sponsored by the UNLV Department of Mathematical Sciences, with funding provided by the National ASA as well. We thank the DMS (ZJ Wu, chair) and ASA for their assistance in pulling this event together.

Morning Session

The morning session focused on regulatory statistics, particularly in the environmental arena. This was prompted by two events. One was a post in June 2016 to the ASA section on Statistics and the Environment (ENVR) discussion site on "Mandatory Statistical Tests in the Code of Federal Regulations", complaining specifically about applying out-dated regulatory language from the late 1980s to a new application area (groundwater monitoring for electric utility coal combustion residuals (CCRs)). This post initiated a good deal of discussion, to the point that a session on this and related issues is in the works for the 2017 Joint Statistical Meetings. The other is that your program co-chair has been heavily involved for some time in another area in which both OSHA and DOE regulation are in a state of flux. Though only a little of the regulation is explicitly statistical, the statistical implications of the proposed regulations demand careful thought.

The session featured Kirk Cameron of MacStat Consulting in Colorado Springs. Dr. Cameron has been involved with environmental regulatory statistics since receiving his PhD from Stanford in 1990. He titled his talk "Influencing Regulatory Policy: Sisyphus in Statistical Garb". Kirk has been very much involved in improving on that late-1980s state of affairs through numerous publications, particularly *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities: Unified Guidance* (an EPA publication which finally



appeared in 2009 after 15+ years of development) along with training courses for regulators and regulated parties. He discussed a number of the technical issues arising in the new CCR regulation. His closing comments on the issue were that there is little hope for sustained progress without statisticians "on the inside", wondering if any were consulted prior to publication of the new CCR rule. He pointed out that many, many regulators have limited statistical under-



standing and/or interest, that to be effective one must understand the "ins and outs" of environmental monitoring, and have some humility, since, in his words, "regulatory policy is much more than science/statistics". (The last seems particularly pertinent these days.) Kirk's presentation will be available on the NV-ASA website.

Charles Davis followed with "On Influencing the Development of Regulation: One Success and One Yetto-be-Determined". He also has been involved with that late-1980s rule; a technique he developed was a basis for an "escape hatch" in the final version of that rule that makes it possible to use techniques that actually make reasonable scientific/statistical sense, though clearly it does not ensure that such will occur. Davis also discussed some of the issues involved in current regulatory development in Industrial Hygiene (Worker Protection) in the nuclear industry. One of these involves statements requiring that certain protections should be required whenever "the concentration" of something exceeds a trigger level. Statisticians immediately realize that we never have a concentration of an environmental contaminant, but rather a distribution of somewhat imprecise measurements of concentrations, which often have interesting temporal and spatial variance components. A first question, then, is should the protections be triggered if the median, or mean, or 95th percentile of the distribution of concentrations exceeds the control level, or if there is any likelihood of any exceedance ever, or when? To further complicate matters, one should implement the protections a priori, before data are available! Communicating these issues with even technically sophisticated regulators, regulation-writers, let alone regulated parties can be a significant challenge.

Afternoon Session

In the afternoon we heard first from Tia Price on "Common Core Myths vs Facts and How It Has Affected Statistics Education". The Common Core State Standards, adopted in October 2010, are known here as the Nevada Academic Content for Standards for English Language Arts and Mathematics (NVACS). Tia took us through a number of the myths ("The standards tell teachers what to teach." "No teachers were involved in

writing" "Adoption ... will dumb down education in Nevada." "Key math topics are missing or appear in the wrong grade."), providing a more reasoned and realistic view of the topics and goals of the NVACS in math and stat. She challenged us with sample end-of-course exam questions that require careful story-problem interpretation along with computational skill. She provided a discussion to be posted to the NV-ASA website.

Then Ashok Singh (UNLV Harrah Hotel College) presented "Perceptions of Nevadans about their Land Transportation System", joint work with Alexander Paz-Cruz, Rebecca Conover, and Hanns de la Fuente-Mella. Data were obtained from a survey conducted online and by interview. Nevada was divided into three regions (NW, NE, and S); questions focused on satisfaction with highway congestion, construction, and safety and on the priorities that should be assigned to various aspects of these. Data presentation was primarily graphical; analyses involved principal components analyses of polychoric correlations of the Likert-scale data. Principal component (PC) scores were compared across the three regions. Briefly, the first PC (representing ~35% of overall variation) seems to be essentially an average of responses to 15 of the 19 questions (excluding those involving funding); its distributions for the NE and NW regions were similar, but different from that for the S region. Other PCs represent at most ~10% of overall variation, and show different and not so extreme regional patterns. The presentation will be on our website. χ

New Leaders of NV-ASA

Chris Tong

Christopher Tong is our new Treasurer. He is a mathematical statistician with the US Department of Agriculture. Based in Sparks, NV, he works remotely for the Center for Veterinary Biologics, located at the National Centers for Animal Health



(NCAH) in Ames, IA. In addition, he collaborates with the National Veterinary Services Laboratory, also located at NCAH. He received a bachelor's degree in physics and math from Knox College, and a PhD in physics and an MS in applied statistics from Purdue University. He was previously employed as a biostatistician in the pharmaceutical industry. He has been an American Statistical Association member for 16 years, and won its Biopharmaceutical Section's best contributed paper award for JSM 2011.

Tia Price

Tia Maria Price, now coordinator of the NV-ASA K-12 Poster Competition, has been an educator for the past 25 years. After teaching high school math for 22 years, she made the transition to secondary math trainer for



the Southern Nevada Regional Professional Development Program (SNRPDP). She assists teachers and school administrators in identifying staff development needs in mathematics as well as planning, organizing, preparing and presenting professional development activities. This includes a comprehensive training workshop for AP Statistics teachers in Southern Nevada. Tia is also a consultant for the National Math and Science Initiative (NMSI), presenting study sessions to students and teachers in the Western part of the US to help prepare them for the AP Statistics exam.

Looking for a Home

Longtime NV-ASA member Ron Titus is offering his collection of statistics journals and books to any interested person (including libraries). The list of publications is extensive and includes *JASA*, *The American Statistician*, *Technometrics*, and *SIAM* publications; the many books are on diverse topics such as regression, multivariate analysis, and statistical consulting. Ron lives in Carson City; if you are interested in any of these (some dating "way back"), please contact Ron at rrtitus@att.net or Deb Stiver at stiver@unr.edu. \(\psi\)



NSC formally launched its Statistics and Applied Mathematics concentration in the Fall of 2016. It has proposed an interdisciplinary data science minor to begin in Fall 2017. κ

֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍

UNLV News

֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍֍

Dr. Lung-Chang (JoJo)
Chien, DrPH, is a new
Assistant Professor in the
Department of Environmental and Occupational
Health, School of
Community Health
Sciences at UNLV. Dr.
Chien came to UNLV from
the Department of



Biostatistics at the University of Texas School of Public Health. His research focuses on spatial vulnerability and geographic disparities on human health; his research topics include socioeconomic deprivation in cancer incidence and mortality, modifying effects of location in geo-survival analysis, the impact of Asian dust storms on children's health, and more.

Long-time NV-ASA member and frequent office holder Hokwon Cho is taking a sabbatical leave this semester. His plans include research trips to UC Santa Barbara and U South Alabama as well as an international conference in Seoul, South Korea.

UNR News **********************

The Department of Mathematics and Statistics at UNR welcomes a new statistician, Dr. Tao Lu. His research interests include longitudinal data analysis, dynamic models, joint models, and Bayesian inference.



Jeff Mortensen is now the acting chair of the department. NV-ASA members

Ania Panorska and Tom Kozubowski of that department have returned from their travels in Europe where they gave several talks on Stochastic Episodes with Light and Heavy Tails and related topics. •



Charles Davis

The afternoon session of our 2014 Fall Symposium featured a round-table discussion prompted by the article "P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume." by Regina Nuzzo (Nature v. 506, 13 February 2014, 150-152). As reported in Newsletter 12(1), we learned that in at least one field (epidemiology) scientific articles that used Pvalues were for a time automatically rejected by the editor! We were (re-)introduced to terms such as Phacking, data torturing, statistician shopping, and publication bias. Sharang Chaudhry discussed the nature of the distributions of P-values. Chad Cross told us about some of the things that actually go on between the time a study is designed and when its results are prepared for submission for publication, as well as how results (including *P*-values and estimated effect sizes) are often interpreted. And Charles Davis discussed the opposite situation, where the customer really wants to avoid statistically significant results; he also introduced ideas of multiple comparisons (how many tests did you run before finding a "significant" result?).

There remains a great deal of interest in these issues in the profession. In late 2015 the national ASA formed a committee to prepare an "ASA Statement on Statistical Significance and *P*-Values", which appeared in the May 2016 issue of *The American Statistician* along with an editorial discussing the history leading up to this formal statement and providing a number of references. The Statement lays out six principles: (1) P-values can indicate how incompatible the data are with a specified statistical model; (2) P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone; (3) scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold; (4) proper inference requires full reporting and transparency; (5) a p-value, or statistical significance, does not measure the size of an effect or the importance of a result; and (6) by itself, a pvalue does not provide a good measure of evidence regarding a model or hypothesis. Each is accompanied by a brief discussion. The final sentence states that "no single index should substitute for scientific reasoning".

The introductory editorial lists some 21 contributions to the discussion, posted as on-line supplements to the published Statement, in addition to nine published references. These present considerable variation in opinion. A fundamental issue occurs in (1) above: in principle the statistical model being tested should be prespecified, as discussed in our 2014 Fall Symposium round-table. For a *p*-value to be valid, not only should the hypothesis and model be specified before experimentation or data collection begin, but all of the data analysis steps should be discussed. And the *p*-value should, again in principle, take the variability inherent in all data-analytic decisions into account. These include

not only model selection but such things as designating outliers and/or data transformations, subdividing the population into "interesting" groups, and so on. These can be challenging even in an experimental or monitoring situation where one has past experience to draw on in identifying appropriate models and anticipated data distribution characteristics.

In other situations, particularly in exploratory or observational studies, it may be very difficult to convincingly argue that one can achieve a convincing level of <u>pre-specification</u>. One area in which this occurs is in the re-analysis of published datasets, sometimes with a view to finding conclusions consistent with an external agenda; see the review of *Doubt Is Their Product: How Industry's Assault On Science Threatens Your Health* in Newsletter 6(1).

Some would argue that such a standard of <u>prespecification</u> is nearly impossible in most scientific studies, and that one should often re-characterize *hypothesis testing* as *hypothesis finding*; that is, finding hypotheses that fit the data currently available which can be tested in future studies. Even then, future studies might be thought of not so much as <u>confirmatory</u> as aimed at helping the hypotheses or models <u>evolve</u> appropriately. Along these lines, please look at the attached position paper by Bert Gunter and Chris Tong. And if this piques your curiosity, look at the on-line supplements they cite; one that catches the eye is "Statistical Tests, *P*-values, Confidence Intervals, and Power: a Guide to Misinterpretations" by Greenland *et al.*

Now we have the forthcoming Symposium on Statistical Inference (Bethesda, MD October 11-13, 2017), entitled "Scientific Method for the 21st Century: A World Beyond p<0.05". The announcement states: "The ASA's statement on p-values and statistical significance addressed many incorrect practices in statistical inference. However, while the statement pointed out what *not* to do, it did not give much guidance on what *to* do. With this in mind, we're now pleased to announce the ... SSI 2017 [which] will focus on specific approaches for advancing scientific methods in the 21st century, considering issues that affect not only research, but research funding, journal practices, career advancement, scientific education, public policy, journalism, and law."

Your comments and discussion will be welcome! *



Only a minority of the people who receive this newsletter are members of the Nevada Chapter of the American Statistical Association (NV-ASA). Dues are

nominal: \$20 per year, \$10 a year for members of the national ASA, and \$2 for student members of the national ASA. There are two ways to join. You can join NV-ASA when you renew your national ASA membership (or join for the first time); this can be done on-line at www.amstat.org/chapters. Otherwise, whether a national ASA member or not, you can join through PayPal on our website or by contacting our Treasurer Chris Tong. Any way you do it, please obtain an information form from our website, complete it, and send it to Dr. Tong at the address listed on the form.

Why should you join? NV-ASA events provide opportunities for networking and contact with other statisticians working in a wide variety of areas in Nevada. But in addition to that, a major reason is that your dues support the outreach activities of the NV-ASA including the K-12 Poster Competition and Career Days. Our financial needs are not great, so long as we all pitch in our modest amounts. *

NV-Chapter Officers & Others

Voting officers are

President: Debra Stiver (2014-2017)
Past President: Debra Stiver (2016-2017)
Northern Vice President: Alejandra Livingston (2017)
Southern Vice President: Charles Davis (2015-2018)
Secretary: Gayle Allenback (2016-2017)
Treasurer: Chris Tong (2017-2018)
Chapter Representative: Anna Panorska (2017-2019)

Also involved are

Webmaster: Alicia Chancellor Hansen

Newsletter co-Editor: Charles Davis Newsletter co-Editor: Hokwon Cho

Co-Editors:

Charles B. Davis Hokwon A. Cho

Typesetting: Hokwon A. Cho

Silver State-istics welcomes news items and letters from members and friends of the NV-ASA on matters of interest to the Chapter and the profession. Manuscript or items can be sent as a Microsoft Word document, PDF, or within an e-mail. Silver State-istics is published by the Nevada Chapter of the American Statistical Association.

© 2017 Nevada Chapter of American Statistical Association

For contact information, go to www.amstat.org/chapters/)



A Response to the ASA Statement on Statistical Significance and P-values

Bert Gunter (Pleasant Hill, CA) and Christopher Tong (Sparks, NV)

The ASA's statement on *p*-values, Wasserstein (2016), and its accompanying discussions are an overdue but welcome response to the "reproducibility crisis" in scientific research. The statement makes a strong case that a major enabler of this crisis has been the widespread use of Null Hypothesis Significance Testing (NHST) and, in particular, the *p*-value as a single numeric summary of the results used to determine scientific validity. Unfortunately, in its single-minded focus on the "misuse and misinterpretation" of *p*-values--and by implication, our profession's non-culpability because we promote "correct" usage--the statement continues to promulgate the fundamentally flawed belief that statistical modeling and inference of some form can objectively characterize scientific uncertainty. We argue here that statistical inference, by which we mean the quantification of the uncertainty in empirical conclusions (as distinguished from, e.g., description and estimation), is fundamentally incompatible with "most" science. Hence, we maintain that insofar as our profession continues to practice and teach the use of inference as an "objective" filter of reproducible science, we remain complicit in enabling the reproducibility crisis.

Our view here is not new; we basically build on the three discussions by Berry (2016), Gelman (2016), and Lew (2016) in particular, but from a somewhat different perspective. We think Gelman & Loken (2014) cogently state the issues: in data analysis, it matters not only what we do but what we might have done. Ergo the critical importance of comprehensive pre-specification of data analytic details for meaningful inference, e.g., as practiced in pivotal Phase III trials in medicine (where even analysis *code* must be provided *before* the trials begin). There is rightful indignation if and when the pre-specified analysis is altered for publication (e.g., http://compare-trials.org/). Almost as diligently, certain subfields of physics require statistical procedures that "must be well understood, well defined, and fixed in advance" (van Dyk, 2014).

However, these are rare scientific examples where the perils of "the garden of forking paths" (Gelman and Loken, 2014) are so assiduously avoided. Rather, we argue that the twin statistical pillars that Tukey (e.g., 1980) called "exploratory" and "confirmatory" are inherently contradictory in scientific research, and that most research falls firmly within the "exploratory" paradigm. Again, contrast this with the clinical trial framework, in which an explicit demarcation is made between "learning" and "confirming" objectives, e.g., Scheiner (1997) and Friedman *et al.* (2010). Statistical inferences for earlier phase clinical trials, and unplanned subgroup analyses of late phase studies, are rarely considered definitive, and are instead fodder for planning later phase or future confirmatory studies, respectively.

In our view, in most scientific research it is desirable to examine empirical results carefully from various perspectives, to look for errors, and to be open to unexpected patterns that may indicate new science. Moreover, as Greenland *et al.* (2016) have noted, the results of statistical inference, be it *p*-values or Bayes factors, are a function not only of the hypotheses (or parameter subspaces) being tested, but of the whole assumed underlying data generation process. In particular, in most scientific research, the strict protocols for randomization and blinding characteristic of clinical trials are rarely adhered to. Not only does experimental bias consequently occur, it is often sought: scientific research programs often need to

build in the flexibility to refine research methodology and measurement protocols over the course of the experimental program. Models (or algorithms) developed as part of the exercise to characterize the data do not and cannot take into account such changes. They are typically overfitted to the final data that result, and do not correspond to any fixed putative population from which the data are randomly sampled. Measures of statistical uncertainty are thus underestimated and of questionable interpretability. See also Deming's (1975) distinction between "analytic" and "enumerative" studies for a closely related discussion.

Unfortunately, as Harrell (2015) pointedly notes, "It is very typical to use the data to make decisions about the form of the model as well as about how the predictors are represented in the model. Then, once a model is developed, the entire modeling process is routinely forgotten, and statistical quantities such as standard errors, P-values, and R^2 are computed as if the resulting model were entirely pre-specified." The same data used to build the model are used to assess the uncertainty of its tests and estimates; the folly of doing so is less obvious there than it is for prediction problems, e.g., Ambroise & McLachlan (2002).

Pre-specification/pre-registration of the statistical analysis plan is often advanced as a cure for inferential ills. Certainly any pre-specification that can limit the data screening, the variables under consideration, the classes of models and fitting methods that will be used and so forth, can improve the replicability of results. But mostly this will be a pale offering that cannot deal with the realities of scientific work. For one thing, key assumptions (randomization, independence, blinding to eliminate bias) still typically will not hold, especially for observational data. But more to the point, pre-specification of the statistical analysis is inconsistent with the exploratory quality of most scientific research. Fine tuning a data analysis to reflect what the data actually say (paraphrasing Tukey), rather than adhering rigidly to prior beliefs and expectations, would seem to be consistent with the principle that science needs to be evidence driven. Model fitting and model criticism may still be perfectly reasonable, but one must be willing to abandon the fantasy that there is objective statistical methodology that can reliably assess the uncertainty in the results. In short, statistical inference requires an unrealistic framework to behave as advertised in exploratory scientific research. Berry's (2016) black-box warning seems to be the best one can say about inferences in this context: "Our study is exploratory and we make no claims for generalizability. Statistical calculations such as p-values and confidence intervals are descriptive only and have no inferential content."

In this light, the one positive role that the ASA statement reserves for *p*-values, measuring "how incompatible the data are with a specified statistical model," is of limited value. The *p*-value (and other outputs of statistical inference) do *not* provide any measure of *model uncertainty* (Chatfield, 1995). A measure of the latter would be difficult to design, given the near impossibility of quantifying the influence of *researcher degrees of freedom* (Simmons et al, 2011) and forking paths (Gelman & Loken, 2015). Thus, the products of statistical inference provide little help in answering the question, am I being fooled by randomness? Technical fixes proposed by the statistical community, e.g., multiplicity adjustments, false discovery rates, penalization and regularization methods, model averaging, and resampling-based estimation of overfitting bias, must also fall short, because they cannot address the basic conceptual tension between exploratory and confirmatory scientific activity.

The use of statistical inference as an adjudicator for scientific legitimacy is a relatively recent phenomenon in science. Freedman (2010) discusses several examples of empirical discoveries from the history of medicine and epidemiology that did not use statistical inference. He argues that "shoe leather," the thoughtful examination of the data closely linked with prior understanding and experience followed by further investigation along multiple lines of evidence, is the way that valid scientific conclusions can be drawn and demonstrated (Freedman, 1991). Naturally this involves good study design and execution, and the evaluation of "many sets of data" (Ehrenberg, 1990). Freedman (1991) wrote that "Generally, replication and prediction of new results provide a harsher and more useful validating regime than statistical testing of many models on one data set. Fewer assumptions are needed, there is less chance of artifact, more kinds of variation can be explored, and alternative explanations can be ruled out." Elsewhere, Freedman (1995) wrote "I wish we could learn to look at the data more directly, without the fictional models and priors. On the same wish list: We should stop pretending to fix bad designs and inadequate measurements by modeling."

We agree. Unfortunately, we believe that statistical insights regarding the design and execution of studies, the descriptive presentation of results, and critical thinking regarding pitfalls of interpretation have yet to fulfill their full potential for improving the scientific enterprise. We think that this is at least in part due to an overemphasis on statistical significance and inferential methodology by our profession. On the other hand, the discussions by Berry (2016) and Lew (2016) suggest one way that inference could be reconceived by researchers. Inferential quantities, such as *p*-values, from exploratory studies are merely elaborate summary statistics, the results of hypothesis *finding* (not hypothesis *testing*) analysis. They should be followed up by (and published along with) confirmatory studies using prespecified analysis plans. Such an approach has been *called pre-registered replication* (Gelman & Loken, 2014; Gelman, 2015), though as the cited authors acknowledge, this approach would not be feasible for use in observational studies, and may be too cumbersome whenever such replication is difficult or impractical. Much exploratory scientific research will need to proceed without the crutch of statistical inference.

We understand that ours is a radical position. But we see the issues as a simple risk/benefit analysis: do the risks of employing inferential methods in science due both to misapplication and fundamental incompatibilities outweigh the benefits? It seems to us that such a judgment should be subject to data---to evidence---and not merely be a matter of opinion. The journal *Basic and Applied Social Psychology* has taken a first step in this direction by banning the use of *p*-values (as well as confidence intervals) in their submissions (Trafimow and Marks, 2015). We need to see how well this works, and whether new alternatives to statistical inference emerge. (We don't pretend to have all the solutions, though we suspect that any solutions that do emerge will be subject matter specific, not as universal as the *p*-value has been.) We hope that other journals in a variety of disciplines will follow their example, so that given sufficient time and replication, a thoughtful analysis of the results can determine how well and in what way scientific research should proceed. And we think the statistics profession should both support and participate in such efforts.

The views expressed here are ours alone, and do not necessarily reflect the policies, views, or opinions of our employers.

REFERENCES

- Ambroise, C. and McLachlan, G. (2002). "Selection bias in gene extraction on the basis of microarray gene-expression data," *Proceedings of the National Academy of Science*, 99, 6562-6566.
- Berry, D. (2016). "P-values are not what they're cracked up to be," on-line supplemental comment to Wasserstein (2016).
- Chatfield, C. (1995). "Model uncertainty, data mining, and statistical inference" (with discussion). *Journal of the Royal Statistical Society A*, 158, 419-466.
- Deming, W.E. (1975). "On probability as a basis for action," The American Statistician, 29, 146-152.
- Ehrenberg, A.C.S. (1990). "A hope for the future of statistics: MSOD," *The American Statistician*, 44, 195-196.
- Freedman, D.A. (1991). "Statistical models and shoe leather," reprinted in Freedman (2010), chapter 3.
- Freedman, D.A. (1995). "Issues in the foundations of statistics: probability and statistical models," reprinted in Freedman (2010), chapter 1.
- Freedman, D.A. (2010). *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*, ed. by Collier, D., Sekhon, J.S., and Stark, P.B. (Cambridge: Cambridge University Press).
- Friedman, L.M., Furberg, C.D., and DeMets, D.L. (2010). *Fundamentals of Clinical Trials*, fourth edition (New York: Springer).
- Gelman, A. (2016). "The problems with *p*-values are not just with *p*-values," on-line supplemental comment to Wasserstein (2016).
- Gelman, A., and Loken, E. (2014). "The statistical crisis in science," American Scientist, 102, 460-465.
- Greenland, S., Senn, S., Rothman, K., Carlin, J., Poole, C., Goodman, S., and Altman, D. (2016). "Statistical Tests, *P*-values, Confidence Intervals, and Power: a Guide to Misinterpretations," online supplemental comment to Wasserstein (2016).
- Harrell, F.E., Jr. (2015). Regression Modeling Strategies, second edition (New York: Springer).
- Lew, M. (2016). "Three inferential questions, two types of *P*-value," on-line supplemental comment to Wasserstein (2016).
- Scheiner, L.B. (1997). "Learning versus confirming in clinical drug development," *Clinical Pharmacology and Therapeutics*, 61, 275-291.
- Simmons, J.P., Nelson, L.D., and Simonsohn, U. (2011). "False-positive psychology: undisclosed flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychological Science*, 22, 1359-1366.
- Trafimow, D., and Marks, M. (2015). "Editorial," Basic and Applied Social Psychology, 37, 1-2.
- Tukey, J.W. (1980). "We need both exploratory and confirmatory," *The American Statistician*, 34, 23-25.
- Van Dyk, D.A. (2014). "The role of statistics in the discovery of the Higgs boson," *Annual Review of Statistics and its Application*; 1, 41-59.
- Wasserstein, R.L., (ed.) (2016). "ASA Statement on Statistical Significance and *P*-Values" (with introductory editorial and on-line discussions), *The American Statistician*, 70, 131-133.