How we might teach and influence interdisciplinary collaborations

Jessica Utts
Professor Emerita
University of California, Irvine
jutts@uci.edu

NISS Webinar, May 23, 2019

Purpose of today's suggestions

- There are many ideas in the lead-in editorial to the special issue of *The American Statistician.*
- I will focus on the most important recommendation, which is eliminating the use of "statistical significance" as defined by a concrete rule.
- My focus is what we can do to help implement the recommendation by working with:
  - Our statistical colleagues
  - Our non-statistical colleagues
  - Journal editors and reviewers
  - Students, including grad and undergrad, statistics and other
  - Teachers of introductory statistics
  - Textbook authors

The recommendation, Wasserstein et al

"The *ASA Statement on P-Values and Statistical Significance* stopped just short of recommending that declarations of "statistical significance" be abandoned. We take that step here. We conclude, based on our review of the articles in this special issue and the broader literature, that it is time to stop using the term "statistically significant" entirely. Nor should variants such as "significantly different," "$p < 0.05$," and "nonsignificant" survive, whether expressed in words, by asterisks in a table, or in some other way. (Page 2)"

Lots of support for this statement; here is one of the most sobering:

"(S)cientists have embraced and even avidly pursued meaningless differences solely because they are statistically significant, and have ignored important effects because they failed to pass the screen of statistical significance…*It is a safe bet that people have suffered or died* because scientists (and editors, regulators, journalists and others) have used significance tests to interpret results, and have consequently failed to identify the most beneficial courses of action." (Rothman, 2016)

Hypothesis testing paradox

- A researcher conducts a test with $n = 100$ and gets these results:
  - $t = \sqrt{100}\left(\frac{\bar{x}-\mu_0}{s}\right) = 2.50$
  - p-value = 0.014, reject null hypothesis
- Just to be sure, the researcher decides to repeat the experiment with $n = 25$

Source: Bob Rosenthal

Hypothesis testing paradox:

- Uh-oh, the results show:
  - $t = \sqrt{25}\left(\frac{\bar{x}-\mu_0}{s}\right)$ = 1.25
  - *p*-value = 0.22, cannot reject null!
  - The effect has disappeared!

- To salvage, researcher decides to combine data:
  - n = 125
  - Finds $t = \sqrt{125}\left(\frac{\bar{x}-\mu_0}{s}\right)$= 2.795, *p*-value = 0.006!
  - The effect is stronger than the first time!

Hypothesis testing paradox:

- Paradox: The 2nd study *alone* did not "replicate" the finding, but when *combined* with 1st study, the effect seems even stronger than 1st study!
- Defining "replication" as getting statistical significance each time, or on the basis of *p*-values, makes no sense! Yet, it's very common practice in many disciplines.

What's going on?

Study	***n***	**Effect size**	$t = \sqrt{n} \times e.s.$	***P*-value**
1	100	**0.25**	2.50	0.014
2	25	**0.25**	1.25	0.22
Combined	125	**0.25**	2.795	0.006

- In all 3 cases the effect size is the *same*, 0.25.
- But the test statistic and *p*-value change based on the sample size, with $t = \sqrt{n} \times$ (effect size).

To be clear…

- We are not advocating getting rid of p-values.
- We are advocating better understanding and more thoughtful interpretation of what they represent.
- We are advocating elimination of the dichotomy of statistical significance, whether at .05 or some other value.

As noted by McShane et al, a better approach is to:

Accept uncertainty and embrace variation in effects: we can learn much (indeed, more) about the world by forsaking the false promise of certainty offered by dichotomous declarations of truth or falsity—binary statements about there being "an effect" or "no effect"—based on some p-value or other statistical threshold being attained.

Why is it so hard to make this change? A sociology lesson

"The basic explanation is neither philosophical nor scientific, but sociologic; everyone uses them. It is the same reason we can use money. When everyone believes in something's value, we can use it for real things; money for food, and P-values for knowledge claims, publication, funding, and promotion. It does not matter if the P-value does not mean what people think it means; it becomes valuable because of what it buys. (Goodman, 2019, p. 27)"

How do we remove the high status of statistical significance as the "coin of the realm?"

Start with statisticians.

Recommendation #1:

Convene a discussion with the other statisticians in your department, school or workplace. Base the discussion on the papers in the *TAS* special issue, especially the editorial. Focus on understanding the big picture problem, which is more likely to engender agreement than the question of how to solve it!

Colleagues who are not statisticians

- There is no “one size fits all” replacement for the current obsession with statistical significance.
- Scientists and statisticians working in a particular discipline need to decide what’s best based on issues like:
  - Is the research mostly cumulative, or are studies unique?
  - Is there expert knowledge that could be incorporated using Bayesian methods?
  - Is the main goal of the research to make decisions, or is it more focused on increasing understanding?
  - Would it be more serious to miss an important finding, or to think something is a problem/solution/difference when it is not?
  - Etc…

Recommendation #2:

If you work with scientists in other disciplines, start a discussion based on the TAS editorial and issues like the ones on the previous slide.

- Ask your collaborators to read the editorial and discuss it with you.
- Find and disseminate papers that demonstrate sound statistical thinking and don't rely on statistical significance (as "best practice" examples).
- Give talks at disciplinary conferences.
- Write an editorial for a journal.
- Contact editors of journals. (See next slides.)

Contacting Journal Editors: Example 1, Nursing Journals

- On May 6, statistician Matthew Hayat and 24 co-signers who collaborate with nursing researchers and educators sent a letter to editors of 18 nursing journals requesting changes in their author submission and publication guidelines. (Next slide.)
- Offered to meet with their editorial boards to discuss.
- Already 5 of the journals have agreed, 1 more is in the process, and 1 has asked for an extended article to publish on the topic. More are considering the suggestions.

The core recommendations in the letter to nursing journals:

"Specifically, in line with the recent effort by the ASA, we recommend including the following manuscript submission requirements:

1) When a *p*-value is reported, state its value regardless of how small or large it may be.

2) Avoid using .05 or any other cutoff for a *p*-value as the basis for a decision about the meaningfulness/ importance of an effect.

3) In reporting a *p*-value, a measure of the effect size should be included, along with a corresponding interval estimate (e.g., confidence interval)."

Contacting Journal Editors: Example 2, Hurlbert, Levine, Utts (paper in TAS)

- Hurlbert sent co-signed letter to over 200 journal editors, grouped by society in same email to encourage discussion.
- Start of letter: *"We write to present a modest proposal for your consideration: that in the instructions to authors for journals under your editorial care it be stated that authors should refrain from use of the phrase 'statistically significant.'"*
- Full letter was 2 pages, reference to the *TAS* special issue and other resources, additional justification.
- Response so far: A dozen recipients have responded that they have our recommendation under discussion by their editors and editorial boards, and four have invited us to submit an op ed, commentary or editorial to their own journal.

What about education? Examples from the *TAS* special issue

Steel, Liermann and Guttorp: "*Beyond Calculations: A Course in Statistical Thinking"* describes a course they teach, intended for senior statistics majors . Quotes:

- "There is a pressing need to move statistical analysis and evidence-based decision making beyond the current system of overtesting and underthinking."
- "Our course is intended as a bridging course. It serves as a bridge between statistical knowledge and successful application of statistics in science."
- "Formally and informally, teaching statistical thinking is, perhaps, our biggest, best, and easiest opportunity to make a positive impact."

Education, example 2: How to conduct a course audit for understanding principles

Maurer, Hudiburgh, Werwinski, and Bailer: *Content Audit for p-Value Principles in Introductory Statistics*

"For a department to perform a comprehensive audit of p-values concepts in their introductory statistics courses, we propose a framework that elicits both intra-departmental and interdepartmental feedback on the current curriculum." They:

- Provide a rubric for evaluating your intro stat courses.
- Provide a framework for focus group discussions with colleagues from across campus.
- Provide resources for others who want to improve understanding of p-value principles in their courses.

Recommendation #3: Help change undergraduate and graduate education

- Investigate the current offerings on your campus to determine whether they cover *understanding* basic principles. (See Maurer et al, *TAS* paper)
- Consider offering a statistical thinking course for majors (see Steel et al, *TAS* paper for excellent suggestions)
- Convene a weekly discussion with statistics graduate students to work through the *TAS* editorial and some of the papers. Make sure none of your students graduate without understanding these ideas!
- Offer to give seminars to graduate students across campus to discuss the ASA p-value statement and the *TAS* editorial.

Quote from Lily Tomlin

"I always wondered why somebody didn't do something about that. Then I realized I AM somebody."

Recommendation #4: Do something!

Questions and comments? Suggestions?

Questions comments anytime: jutts@uci.edu