Discussion of The Introductory Statistics Course at a Crossroads: Key Next Steps for the Future

Jeff Witmer
Oberlin College, 10 N. Professor Street, Oberlin, OH 44074

Abstract
I respond to the ideas presented by Nicole Lazar (p-values are a problem), Nick Horton (use computers to explore the multivariate world), and Kevin Cummiskey (casual reasoning is important in the multivariate world).

Key Words: Education, GAISE guidelines

1. Nicole Lazar: What to teach (and not to teach) about p-values

Lazar claims that p-values are a problem and that statistics educators should do something about that. I agree. (How could I not?) The ATOM(ic) principles that Lazar promotes are valuable, especially the Acceptance of uncertainty. I'll add that reporting p-values to several decimal places implies that they are much more useful than they really are. I'm very happy to join Lazar in endorsing simulation-based inference.

We should be estimating effects and reporting effect sizes. But I'm going to mainly talk about terminology.

I like to emphasize that "statistical *significance*" is a very poor term and that it misleads not only the public but some hypothesis testers as well. For example, in a recent submission to the *Journal of Statistics Education* I read "Despite different cultural and geographical differences, we found that the three models as a whole do not significantly differ." My question is "Does this mean they do not *meaningfully* differ?"

In a recent *JAMA* article I read "Universal masking at MGB [Mass General Brigham] was associated with a significantly lower rate of SARS-CoV-2 positivity among HCWs [health care workers]." Does this mean that the rate is *importantly* lower?

Do students understand what "statistically significant" means? Here is a sentence from a recent final exam:

> "Since this p-value is less than 0.05, that means that H0 is rejected and there is a significant difference in degree completion status between men and women."

Does this student understand that the small p-value means that the men versus women difference in the sample was beyond what chance variation would predict, but it might be so small that no one would really care about it?

I prefer to use the terminology "statistically *discernible*" and to say that the test discerned a difference between the population means; that the sample means were so far apart that we can infer that the population means are actually different.

Calling the observed (sample) difference "statistically significant" obscures what is happening.

Here is another example, from https://chicago.suntimes.com/2019/10/25/20932513/stroke-heart-attack-risk- blood-pressure-medicine-before-bed:

> Those who took their medicine at night "showed significantly lower" rates of issues resulting from high blood pressure.

My question is, does this mean "discernibly lower – we had a small p-value" or "importantly lower – this makes a real difference, and my doctor should care"?

I can't tell.

I could go on, but the point is that STAT 101 students should *not* be taught that what statisticians do is determine whether or not a p-value is small, and if it is then they tell the world that the result is "significant."

2. Nick Horton: The role of computing at the core of a modern introductory statistics course

Horton claims that the world is multivariate and we can explore it with computers and good software. I agree. (How could I not?) As Terry Speed wrote (in 1986, at ICOTS2): "Statistics is no closer to mathematics than cooking is to chemistry....statistics and cooking are as much arts as they are science."

Learning the component parts isn't enough. We need to always ask What question are we trying to answer? Students need to build models and to learn concepts.

I'm very happy to join Horton in endorsing RMarkdown.

I should mention that I was a more popular teacher back when I spent a lot of time on formulas. (I remember reading the Instructor's Manual for Freedman, Pisani, and Purves and the note reporting that mathematically weak students complained about needing to learn formulas, but they complained more loudly when those formulas were taken away, so that the students had to really think and understand ideas, rather than memorize and learn how to plug in numbers.) I might be less popular, but today my students learn more, and leave my courses much better prepared.

3. Kevin Cummiskey: Multivariate thinking, confounding, and causal inference in introductory statistics

Cummiskey claims that multivariate thinking is important and can help us understand what causes what. I agree. (How could I not?) But statisticians (particularly statistics educators) have steered away from talking about causation, unless the data are from a randomized experiment. This is a mistake.

And it isn't just statisticians who hesitate. From *Perspectives on Psychological Science* (summer, 2020): *"Causal inference is a central goal of research. However, most psychologists refrain from explicitly addressing causal research questions and avoid drawing causal inference on the basis of nonexperimental evidence."*

As educators, we should show repeated examples of thinking about causal relationships. Causal diagrams are very helpful – but first we have to agree that *talking about and teaching causal inference is an important STAT 101 topic*. I like the idea of telling students that when we see an association it might be due to (a) causation, or (b) mere association, or (c) randomness, i.e., chance variation in the data when there is no true association in the larger population.

4. Conclusion

All three speakers presented good ideas that are the result of deep thinking about what we do as statistics educators. Horton and Cummiskey drew on the revised GAISE guidelines (Guidelines for Assessment and Instruction in Statistics Education) and Lazar drew on the 2019 paper "Moving to a World Beyond '$p < 0.05$'" that has generated much reaction. I hope that the three talks here generate similar reaction and that educators follow the advice of these three speakers.

References

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Rowell, G., Velleman, P., Witmer, J., and Wood, B. (2016), "Guidelines for Assessment and Instruction in Statistics Education: College Report 2016," American Statistical Association.

Wasserstein, R., Schirm, A., and Lazar, N. (2019). "Moving to a World Beyond '$p < 0.05$.'" *The American Statistician*. 73 (sup1): 1 – 19.