Experiencing a Paradigm Shift: Teaching Statistics through Simulation-based Inference

Dave Klanderman, Mandi Maxwell
Department of Mathematics, Trinity Christian College, Palos Heights, IL 60463, USA
Dave.Klanderman@trnty.edu, Mandi.Maxwell@trnty.edu
and Nathan Tintle, Department of Mathematics, Computer Science and Statistics, Dordt College, Sioux Center, IA, 51250, USA, Nathan.Tintle@dordt.edu

Abstract

For decades, introductory statistics has been taught as an application of formulas, making use of normal and other distributions, and relying heavily on algebraic skills of students, in short, emphasizing mathematical thinking. More recently, several textbook author teams have published statistics textbooks that place an increased emphasis on simulation and randomization methods as the way to motivate statistical reasoning (e.g., inference) leading to a decreased emphasis on the algebraic manipulation of formulas and theory-based approximations to sampling distributions (e.g., [3]; [7]). This paper describes simulation-based inference curricula more fully, reports on the necessary steps towards the implementation of such an approach, and provides both qualitative and quantitative comparisons of this new pedagogical approach with a more traditional approach. Appropriate justification of the approach to teaching and learning statistics is also provided, along with providing an overview of recent trends to shift to this approach in statistics courses taught at the high school, junior college, and university levels across North America, including a number of Christian colleges and universities affiliated with the ACMS.

Introduction

If you have taught an algebra-based introductory statistics course (e.g., the AP statistics equivalent course) during the past few decades, then perhaps the following description is familiar. Topics normally proceed from descriptive statistics to probability theory, from sampling distributions and the Central Limit Theorem to inferential statistics focused on hypothesis tests and confidence intervals. The variety of different tests and intervals normally includes a single mean, two means, ANOVA, a single proportion, two proportions, Chi-Square, and correlation/regression. Typically, these courses make extensive use of formulas, tables of values for t, z, χ^2, and F distributions, graphing calculators, and/or any of a wide array of statistical packages (e.g., SPSS, Minitab, Excel, R). In short, this standard introductory statistics course consists of a collection of theory-based techniques and is often taught from a perspective of an application of mathematical formulas rather than as a course in statistical thinking.

A Challenge to this Paradigm

George Cobb [2] offers a sharp critique of this familiar version of the introductory statistics course. Among other issues, he notes that many of these theory-based techniques have validity conditions that are not satisfied in many real data contexts. Furthermore, Cobb notes that some of these techniques have been modified over the years to overcome some of the challenges posed by these validity conditions. As an example, the “Plus

4" method for computing confidence intervals for a single proportion was developed in the 1990s as a way to overcome the need for a minimum number of successes and failures in the sample data. Cobb refers to examples such as this one as a "cathedral of tweaks" and a futile effort to eliminate this shortcoming of theory-based methods. In these situations, Cobb posits that, at best, students will memorize technical rules, losing sight of the big picture of how to think statistically.

As an alternative, Cobb and a larger team of statistical educators and researchers proposed a new approach to the teaching, learning, and doing of inferential statistics. This paradigm makes extensive use of simulation, both tactile and computer-based via applets developed by Beth Chance and Allan Rossman (see [1] for more details on the rationale for these applets. Access them at http://www.rossmanchance.com/ISIapplets.html). The applets use a combination of resampling and randomizations to produce large scale simulations based upon the sample data. An experimental p-value based upon these simulations can be easily computed, and this value will be approximately equal to the p-value using the comparable theory-based approach, assuming that the validity conditions are satisfied. Interestingly, when these two p-values do not match, the fault lies with the validity conditions but the simulated p-value can still be used. In the past few years, at least two textbooks have been written that make extensive use of simulation and randomization, including appropriate applets (cf., [3]; [7]).

Highlights of the New Paradigm for Teaching and Learning Statistics

By designing an introductory statistics course using the resources described above the teaching and learning paradigm may differ in several important ways. First, an overview of the scientific method as applied to statistical analysis is possible from the beginning of the course using a discussion of inference on a single proportion as an initial case study with inferential statistics following closely thereafter. Roy and her colleagues [5] make the pedagogical case for introducing p-values in this manner during the first week of such a course. Essentially, students have good intuition in this situation and can relatively easily understand the core logic of inference of simulation with a single proportion without getting stuck in technical definitions (e.g., sampling distribution; binomial distribution; normal approximation to the binomial, etc.) The remainder of the course allows students to deepen their understanding of the overall process. P-values and overarching themes such as significance, estimation, generalization, and causation are investigated. Descriptive statistics, including means, medians, standard deviations, five number summaries, dotplots, boxplots, correlation, and least squares regression lines, emerge in a "just in time" manner as the concepts are needed to perform related statistical tests. Similarly, those probability concepts necessary for understanding the inferential statistics are introduced as the need arises.

A brief example from the [7] textbook will illustrate the general approach. Like essentially all other examples and explorations in the book, this example is taken from actual research. Two dolphins, Buzz and Doris, were placed in a pool but separated by a curtain. Doris was instructed to communicate with Buzz who, in turn, pressed one of two buttons. A correct result by Buzz was followed by fish being fed to both dolphins. In a total of 16 attempts, Buzz answered correctly 15 times. Initially, students are asked whether this data constitutes evidence that the dolphins are communicating, and most of the students agree that this is the case. When asked how they would justify this finding, the discussion eventually leads to the need to simulate what would happen if Buzz was guessing, by using an ordinary coin flipped a total of 16 times, repeatedly. A class set of results normally suffices to convince students that an outcome of 15 (or more) successes is an extremely rare event, assuming that Buzz is merely guessing. At this point, the one proportion applet is introduced and students use their own smart phones or laptops to simulate Buzz guessing for a total of 1000 or 10,000 sets of 16 flips, noting the proportion of repetitions in which 15 or more successes are observed (see Figure 1).

Notice that students must eventually learn to interpret this simulated p-value as the likelihood of obtaining 15 or more successes out of 16 attempts assuming that the null hypothesis (i.e., Buzz is guessing) or chance

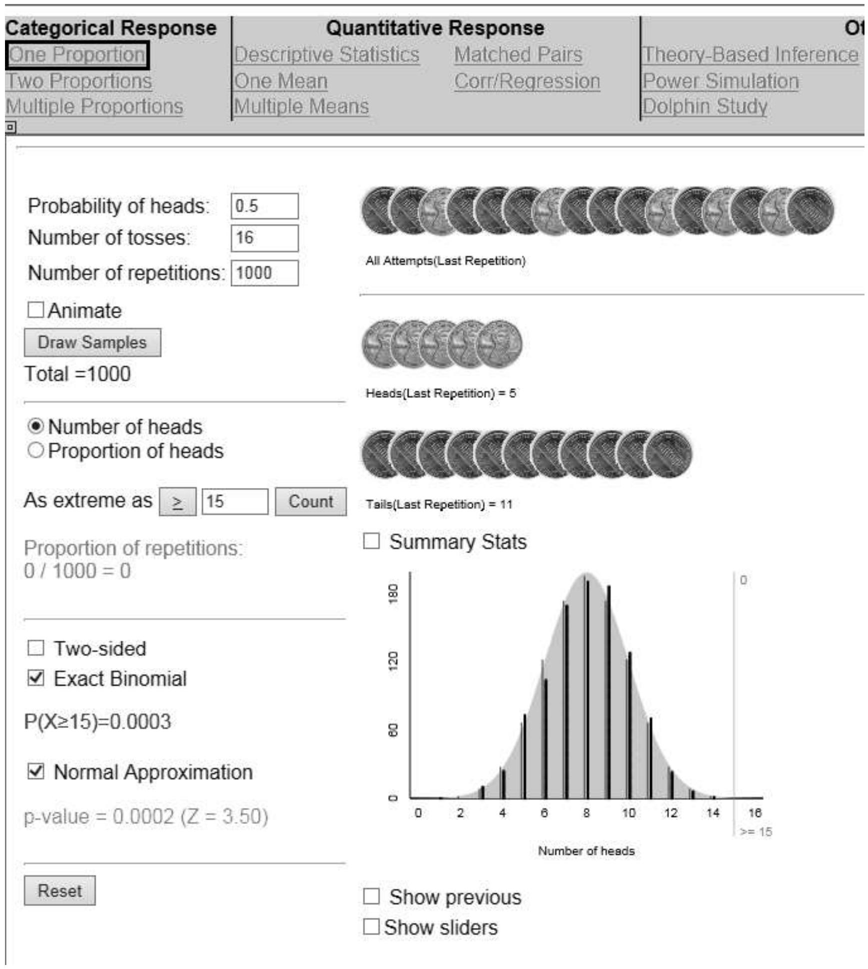

Figure 1: One Proportion applet to model initial Buzz and Doris Example

model is true. An extension of this example involves a follow-up experiment in which Buzz was successful only 16 out of 28 times. This time the applet provides a nice visual image for the simulated p-value and students readily conclude that these data do not provide strong evidence that the dolphins are communicating. As a final note on this example, the researchers later acknowledged that they stopped feeding the dolphins at some point during this latter study so the results are hardly surprising (see Figure 2). Note that the applet also provides the exact p-value using the binomial distribution along with an estimate using the normal distribution which is not accurate due to the discrete nature of the data. Other topics in the introductory statistics curriculum can be similarly addressed using simulation, randomization, and corresponding applets.

More Applets and Potential Uses in an Advanced Probability and Statistics Course

When an advanced course in probability and statistics is offered at Trinity Christian College, the need to develop probability theory, probability distributions, expected values, and variances, limits the time spent on inferential statistics to the final five or six weeks of the course. For this reason, there is typically not sufficient time to develop the Buzz and Doris and similar one proportions tests in great detail. However, other topics normally covered in this portion of the course can be enhanced by using the applets and demonstrating how simulation and randomization provide an excellent approximation to the theory-based p-values.

One such example is the matched pairs test. Although this test often poses some initial confusion for students,

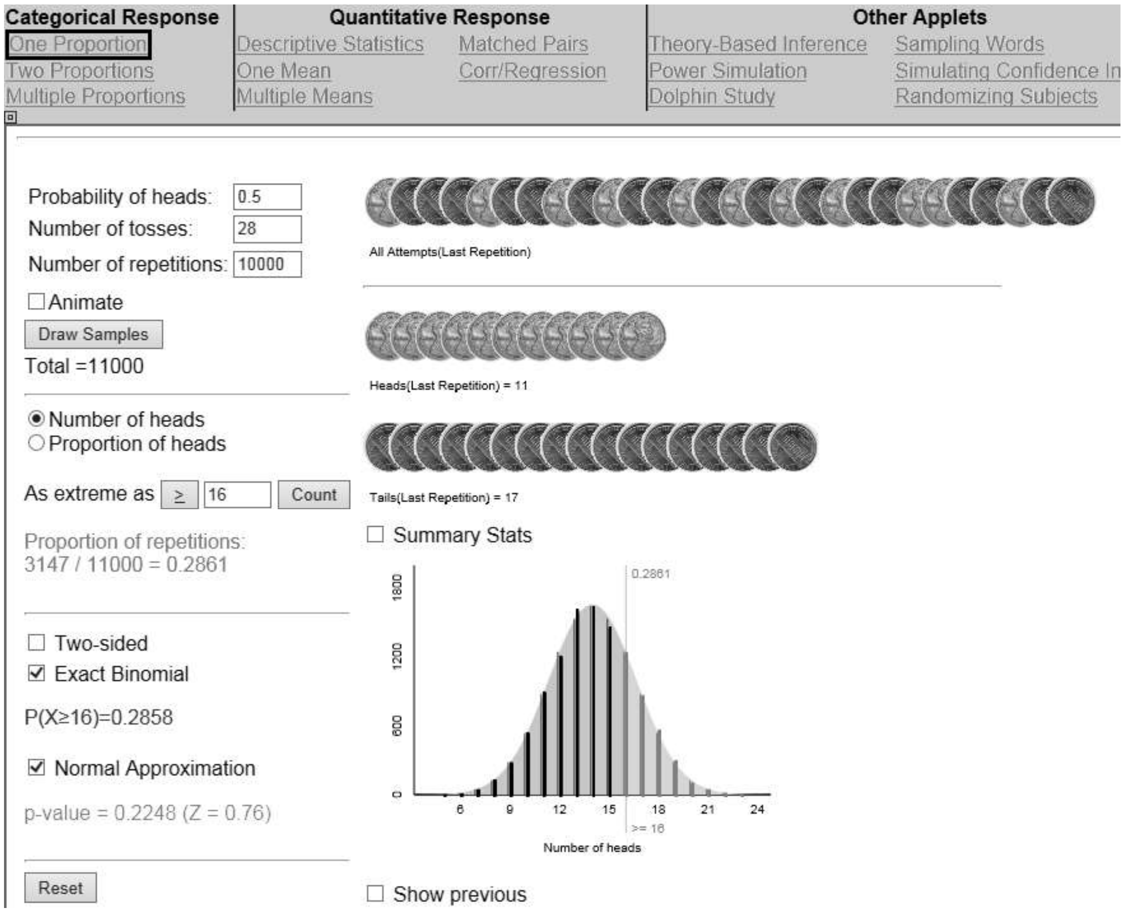

Figure 2: One Proportion applet to model follow-up Buzz and Doris Example

the applet makes the reason for the pairing of the data values clear visually and highlights how this method, when applied correctly, dramatically reduces the variability in the sample. Simulation can then be done to produce an approximate p-value (see Figure 3). It is important to note that applets are also provided to perform the theory based test for one mean (as in the case of the matched pairs test), two means, one proportion, and two proportions. For other tests, including multiple proportions, multiple means, and the slope of the regression line, the theory-based method is included as an option within the simulation applet. In the Tintle et al. text, [7] sections at the end of each chapter provide an overview of the theory-based method, including a discussion of the necessary validity conditions.

A second example is correlation and regression. In this area, the applet provides a helpful interpretation of the Least Squares Regression Line. It allows students to experiment with many possible lines through the data, and it provides a comparison of the total squared error in each case, both numerically and graphically (see Figure 4). As before, simulation can be performed to produce an approximate p-value.

Finally, students often find it difficult to correctly interpret the meaning of confidence intervals. Typically, the student will claim that there is a 95% probability that the actual mean will fall inside a 95% confidence interval for μ when, in fact, this is a nonsensical statement since the interval either did or did not succeed in trapping the parameter value, and we typically do not know because μ is unknown. The Simulating Confidence Intervals applet provides an opportunity for students to construct 100 different 95% confidence intervals for the mean. Typically, all but around 5 or so of these intervals will include the correct value (which must be known for this illustration) (see Figure 5).

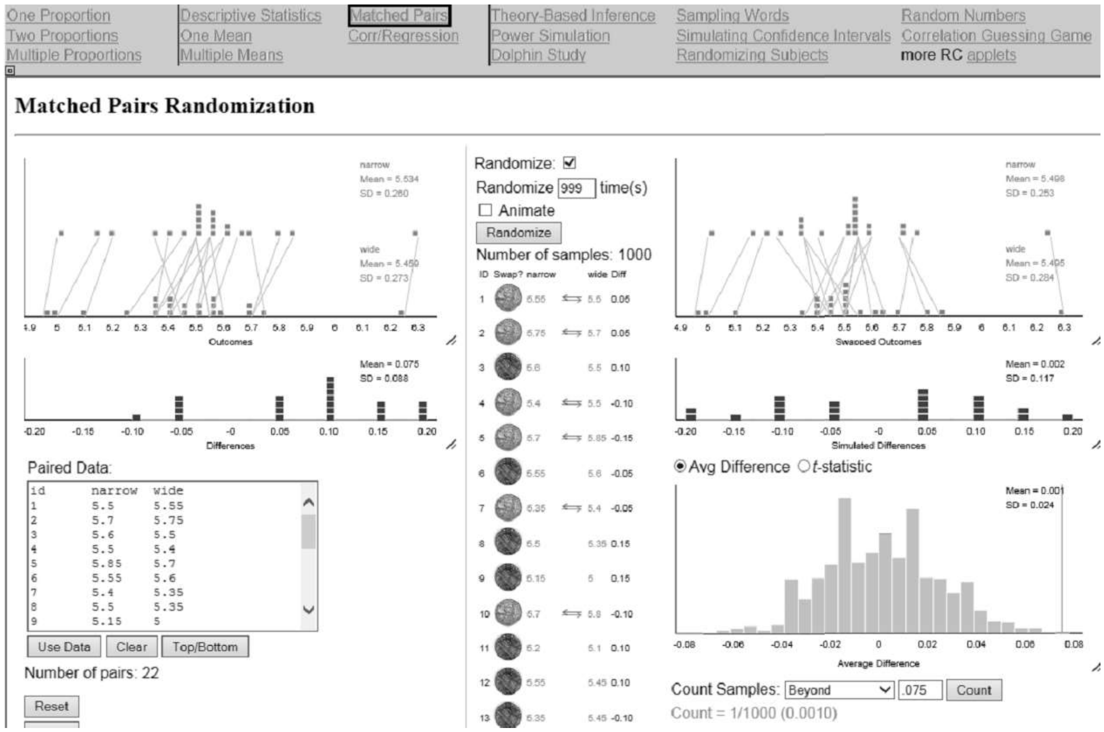

Figure 3: Matched Pairs Applet

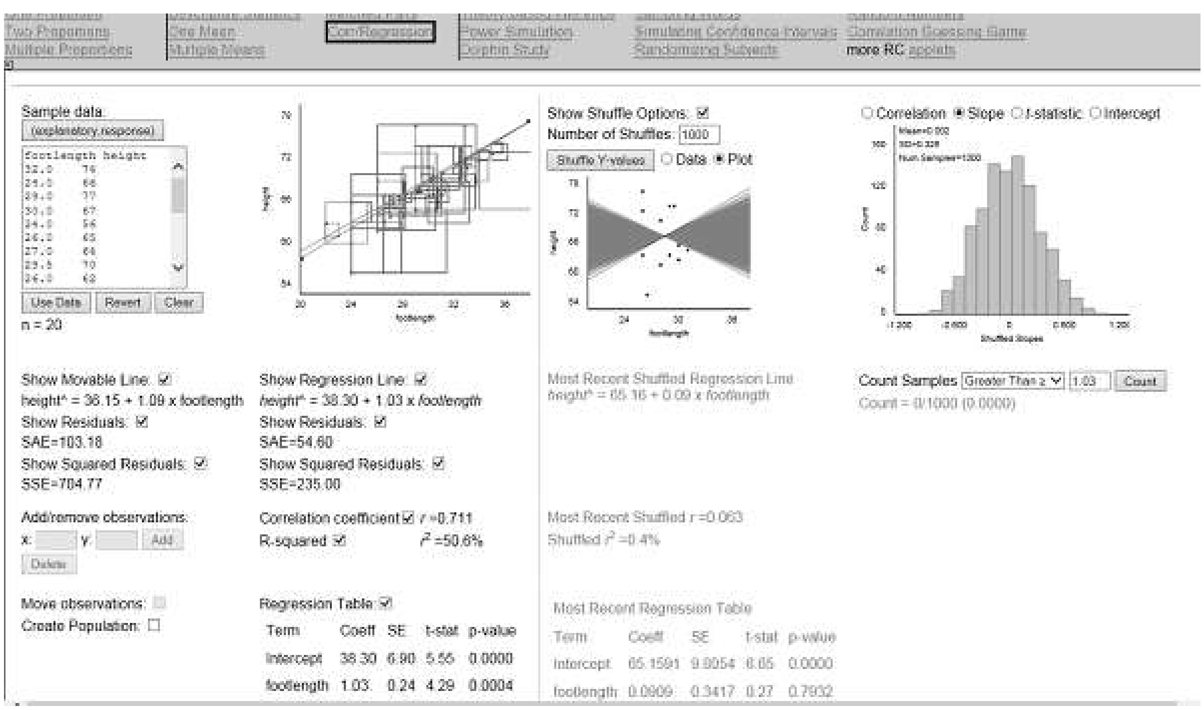

Figure 4: Correlation and Regression Applet

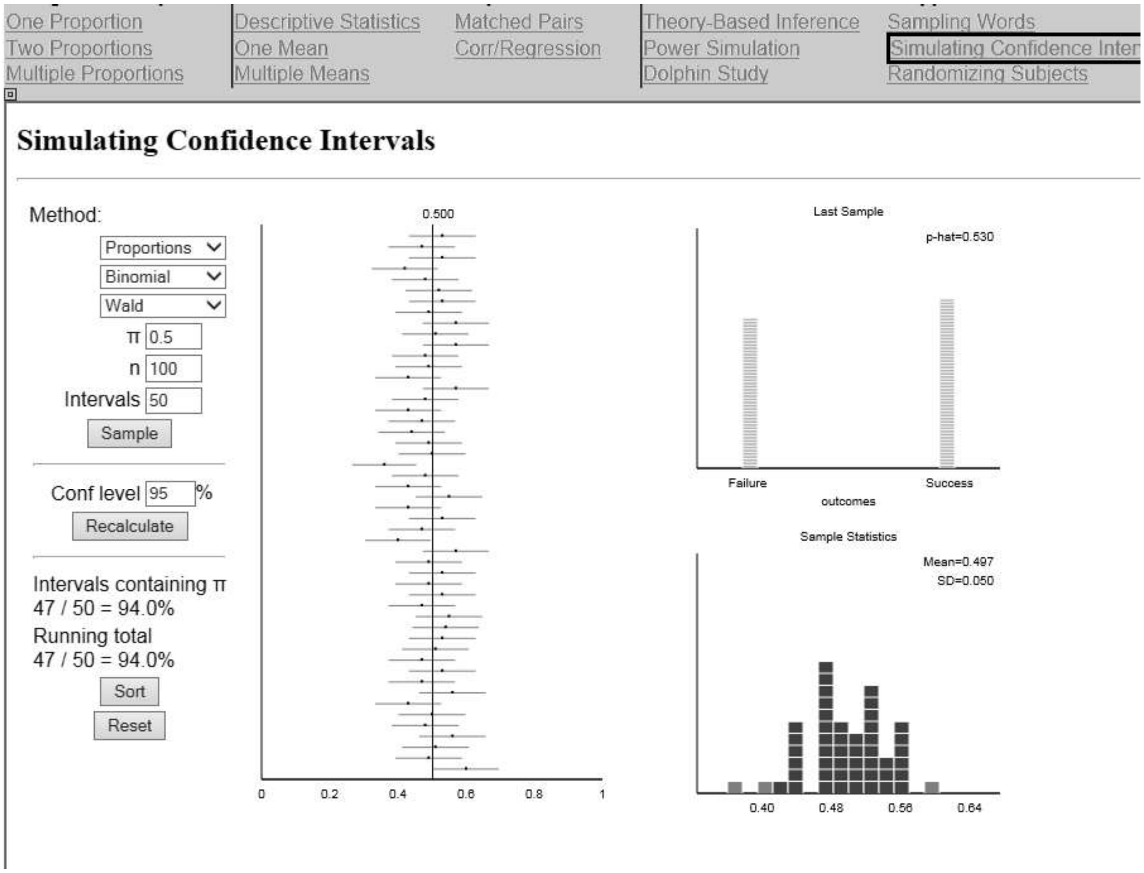

Figure 5: Simulating Confidence Intervals Applet

Results of Teaching and Learning Introductory Statistics Using Simulation-Based Methods

Tintle and his colleagues [11] report on quantitative results comparing conceptual knowledge of statistical concepts by students taking courses using this new paradigm and those taught using the older and more familiar paradigm. A 40-item multiple choice instrument, based upon the Comprehensive Assessment of Outcomes in Statistics (C.A.O.S.) was administered at three different times: during the first week of the course, during the last week of the course, and four months following the completion of the course. For 33 of these 40 items, both groups of students performed similarly. However, for 6 of the remaining 7 items, the students who learned using simulation-based methods outperformed their counterparts in the more familiar statistics courses. Not surprisingly, one of these 6 items assessed a student's knowledge of a p-value. These findings have now been replicated at multiple institutions [9]. Perhaps more importantly, there was less of a decline in performance four months following the course for the students in the simulation-based course [10]. That is, these students retained more of their accumulated conceptual knowledge of statistics as they enrolled in courses in their major that required them to apply this understanding. Swanson, VanderStoep, and Tintle [6] document similar findings related to student attitudes towards statistics and the learning of new statistical concepts.

The first author of this paper used a well-known traditional text [4] during the fall of 2013 before switching to a text using simulation and randomization [7] during the three subsequent semesters. Overall, the percentage of students earning grades of C or better in the course was similar for both approaches. Like in the traditional course, some students struggled in the simulation-based course. However, the focus of these struggles were less algebraic and arithmetic in nature and more connected to the statistical concepts such as causation and generalization. Equally important, the use of real data and explorations promoted a greater level of student engagement in the simulation-based course.

Conclusion

If you are currently teaching introductory statistics using the old paradigm, then we encourage you to try a simulation-based approach. You may also learn more by reading Cobb's article [2], or by reading a more recent article related to the impact of simulation ideas on the entire undergraduate statistics curriculum [8], or by visiting the simulation-based inference blog where numerous statistics educators post on their experiences using simulation-based inference in their classes. Although it would be helpful to augment your current course with the use of some of these freely available applets, you will derive the greatest benefit and personal satisfaction if you accept the challenge to join the new teaching and learning paradigm. Enrolling in a 2-4 day intensive workshop on these approaches would also greatly enhance your chance of success in this effort. Please contact any of the authors for additional information.

References

[1] B. Chance, D. Ben-Zvi, J. Garfield, and E. Medina. The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, 1(1), 2007.

[2] G. Cobb. The introductory statistics course: A Ptolemaic curriculum. *Technology Innovations in Statistics Education*, 1(1), 2007.

[3] R. Lock, P. F. Lock, K. L. Morgan, E. Lock, and D. Lock. *Statistics: Unlocking the Power of Data*. John Wiley & Sons, Inc., Hoboken, NJ, 2012.

[4] D. Moore. *The Basic Practice of Statistics*. W.H. Freeman and Company, New York, third edition, 2004.

[5] S. Roy, A. Rossman, B. Chance, G. Cobb, J. VanderStoep, N. Tintle, and T. Swanson. Using simulation/randomization to introduce p-value in week 1. In K. Makar, B. deSousa, and R. Gould, editors, *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*, 2014.

[6] T. Swanson, J. VanderStoep, and N. Tintle. Student attitudes toward statistics from a randomization-based curriculum. In K. Makar, B. deSousa, and R. Gould, editors, *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*, 2014.

[7] N. Tintle, B. Chance, G. Cobb, A. Rossman, S. Roy, T. Swanson, and J. VanderStoep. *Introduction to Statistical Investigations*. John Wiley & Sons, Inc., Hoboken, NJ, preliminary edition, 2015.

[8] N. Tintle, B. Chance, G. Cobb, S. Roy, T. Swanson, and J. VanderStoep. Combating anti-statistical thinking through the use of simulation-based methods throughout the undergraduate curriculum. `http://math.hope.edu/isi/presentations/white_paper_sim_inf_thru_curriculum.pdf`, 2014.

[9] N. Tintle, A. Rogers, B. Chance, G. Cobb, A. Rossman, S. Roy, T. Swanson, and J. VanderStoep. Quantitative evidence for the use of simulation and randomization in the introductory statistics course. In K. Makar, B. deSousa, and R. Gould, editors, *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*, 2014.

[10] N. Tintle, K. Topliff, J. VanderStoep, V.-L. Holmes, and T. Swanson. Retention of statistical concepts in a preliminary randomization based introductory statistics curriculum. *Statistics Education Research Journal*, 11(1), May 2012. `http://www.stat.auckland.ac.nz/~iase/serj/SERJ11(1)_Tintle.pdf`.

[11] N. Tintle, J. VanderStoep, V.-L. Holmes, B. Quisenberry, and T. Swanson. Development and assessment of a preliminary randomization-based introductory statistics curriculum. *Journal of Statistics Education*, 19(1), March 2011. `http://www.amstat.org/publications/jse/v19n1/tintle.pdf`.