Challenging the State of the Art in Post-Introductory Statistics: Preparation, Concepts, and Pedagogy

Nathan L. Tintle[1*], Beth Chance[2], George Cobb[4], Allan Rossman[3], Soma Roy[3], Todd Swanson[5], Jill VanderStoep[5]
[1]Dordt College, Sioux Center, Iowa, USA
[2]California Polytechnic University, San Luis Obispo, California, USA
[3]Mt. Holyoke College, South Hadley, Massachusetts, USA
[4]Hope College, Holland, Michigan, USA
*Corresponding author: Nathan Tintle, e-mail: nathan.tintle@dordt.edu

Abstract

The demands for a statistically literate society are increasing, and the introductory statistics course ("Stat 101") remains the primary venue for learning statistics for the majority of high school and undergraduate students. After three decades of very fruitful activity in the areas of pedagogy and assessment, but with comparatively little pressure for rethinking the content of this course, the statistics education community has recently turned its attention to use of randomization-based methods to illustrate core concepts of statistical inference. This new focus not only presents an opportunity to address documented shortcomings in the standard Stat 101 course (for example, improving students' reasoning about inference), but provides an impetus for re-thinking the timing of the introduction of multivariable statistical methods (for example, multiple regression and general linear models). Multivariable methods dominate modern statistical practice but are rarely seen in the introductory course. Instead these methods have been, traditionally, relegated to second courses in statistics for students with a background in calculus and linear algebra. Recently, curricula have been developed to bring multivariable content to students who have only taken a Stat 101 course. However, these courses tend to focus on models and model-building as an end in itself. We have developed a preliminary version of an integrated one to two semester curriculum which introduces students to the core-logic of statistical inference through randomization-methods, and then introduces students to approaches for protecting against confounding and variability through multivariable statistical design and analysis techniques. The course has been developed by putting primary emphasis on the development of students' conceptual understanding in an intuitive, cyclical, active-learning pedagogy, while continuing to emphasize the overall process of statistical investigations, from asking questions and collecting data through making inferences and drawing conclusions. The curriculum successfully introduces introductory statistics students to multivariable techniques in their first or second course.

Keywords: Variability; Stat 101; Randomization-based; Second course

1. Introduction

During the past 30 years, the advent of the use of computers in statistics and resulting discussions about possibilities for teaching statistics ushered in an era of serious scientific inquiry about how and what students learn about statistics in their courses (Garfield 1995, Garfield and Ben-Zvi 2007), through more systematic investigation of student reasoning and through the use of sophisticated assessment instruments (e.g., ARTIST 2005). This drive towards improved statistical education led first to the Teaching Statistics report (Cobb 1992), and later the Guidelines for Assessment and Instruction in Statistics Education (GAISE, 2005) which emphasized focusing on statistical literacy and statistical thinking, the use of real data, conceptual understanding, active learning, the use of appropriate technology, and improved use of assessments.

After three decades of very fruitful activity in the areas of pedagogy and assessment, but with comparatively little pressure for a complete overhaul in content, the statistics education community has recently been exploring the need for new content and focus in the introductory statistics course (Stat 101). Randomization-based and computer-intensive materials for the first course constitute one major response to this need (e.g., Lock et al., Tintle et al. 2013). The development of new materials for a second statistics course constitutes another response (e.g., Cannon et al. 2013, Kuiper and Sklar 2013). The growing number of proposed alternatives to the traditional course in mathematical statistics constitutes yet a third (e.g., Chihara and Hesterberg 2013; Chance and Rossman, 2013, for more see Cobb (2011)), and there are important others as well (e.g., Nolan and Temple-Lang 2007, Kaplan 2012). The push for different content and focus is driven in large part by external changes which have made drawing conclusions from data ubiquitous in our society. The growing multidisciplinary data tidal wave ("big data") necessitates a more statistically literate society lest its citizens be forced to adopt an all or nothing approach to trusting or mistrusting data-driven decisions.

Randomization approaches (permutation tests, bootstrapping and simulation), which may allow students to more quickly understand and appreciate the core logic of statistical inference (Cobb 2007), are gaining momentum as a reasonable approach for Stat 101, as evidenced by proof-of-concept materials (Lock et al. 2013, Tintle et al. 2013, Tabor and Franklin 2011, among others) and promising early assessment results (Tintle et al. 2011, Tintle et al. 2012). Thus, it could be argued that the motivation behind randomization approaches in Stat 101 is to improve student understanding of the logic of inference through an alternative content approach that is conducive to GAISE pedagogy.

With some momentum and focus for content revision ongoing for Stat 101 via randomization approaches, the second course in statistics is now receiving attention, as the content and focus of the curriculum are dissected. Historically, a student's second course in statistics has been either regression or ANOVA/Design of Experiments, with the vast majority of such courses utilizing calculus and linear algebra concepts, substantially limiting the audience that would take such courses. Increasingly, however, there is a move to consider how methods that now dominate statistical practice (e.g., multiple regression; Switzer and Horton 2007; Lohr 2011) can be brought to the rapidly growing Stat 101 audience—most of whom will never take calculus or linear algebra. One of the most popular books for second courses (Stat 201) has been <u>The statistical sleuth</u> (Ramsey and Schafer, 2013), with the recent addition of two more alternatives, <u>Stat 2: Building models for a world of data</u> (Cannon et al. 2013) and <u>Practicing Statistics: Guided Investigations for the Second Course</u> (Kuiper and Sklar, 2011). The common theme across all of the books is the building of models (e.g., ANOVA; multiple regression; general linear models; ANCOVA)—in fact, the majority of content of these books is devoted to learning the appropriateness and nuances of a variety of multivariable models. One could argue that the approach of these recent Stat 201 courses is a focus on the concept of what a statistical model is and how it should be used.

We applaud the efforts to date which have provided innovative texts so that second courses in statistics can be taught to the multitudes of post-Stat 101 students. We believe that serious discussion about the focus and concepts taught in second courses in Statistics may bring about changes to the curriculum that would more broadly embrace GAISE guidelines and widen the audience of potential post Stat 101 students. Here, we wish to posit that the goal of a second course may not be statistical models themselves, but instead should be the concept and strategies that motivate and structure the modeling process. In the following sections, we will (1) Lay out an alternative conceptual goal of a second course in Statistics—to enhance students' understanding of confounding and variability through an approach which is not model-centered, but is instead, meaning-centered, (2) Present an outline for a curriculum that places conceptual understanding of variability and confounding at the center, (3) Describe preliminary results from implementation of such a curriculum and (4) Conclude with a discussion of potential next steps for the field in discussing second courses in statistics.

2. An alternative approach: Not model-centered but meaning-centered

Much as Cobb has argued (2007) that the goal of Stat 101 should not be the techniques for doing statistics inference, but should be the core logic of inference (an approach facilitated greatly by randomization approaches), we suggest that the goal of a second course in statistics should not be the models for conducting multivariable statistical inference, but should instead be the concepts that underlie the motivation and utility of such models: namely, confounding and variability.

In a first course which focuses on the core logic of inference (e.g., our randomization-text, Tintle et al. 2013), a major focus is on whether observed data is typical or atypical given certain (null) hypotheses about the process, population, or experiment which generated the data, with randomization and simulation can be used to evaluate strength of evidence against the (null) hypothesis,. Thus, by focusing on the core logic of inference we are often focus on chance variability. Like most Stat 101 courses, ours does not frequently discuss multivariable datasets, instead exploring how chance variability can be evaluated across a variety of tests of one- and two-variables.

In multivariable datasets, confounding between measured (and unmeasured) variables becomes increasingly prominent, with a host of statistical techniques (models) attempting to evaluate the extent of the confounding. Relatedly, co-variation (or lack thereof) between variables and understanding how to minimize co-variation through design and analysis strategies lies at the heart of multivariable data analysis practice.

We suggest that students should better understand how confounding and variability are two major obstacles to learning from data, and that design and analysis decisions can limit the effect of these obstacles. In particular, we argue that enhancing student understanding of these ideas will result in a deeper student understanding of how we can strategize to maximize learning from data and how we can appropriately interpret data. And so, *modeling*, the typical theme of the second course, assists in achieving better student understanding of these concepts, but is not an end in itself.

3. An outline for a curriculum that places conceptual understanding of confounding and variability at the center

If the conceptual goal of a second course in statistics is confounding and variability, the motivation for multivariable models, how does this impact curriculum design decisions? To answer this question, consider that current post Stat 101 courses typically start by focusing on the unity of two variable approaches (e.g., two-sample *t*-test, ANOVA and regression), and then proceed to let the general linear model drive the remaining content, yielding chapters on ANCOVA (analysis of covariance—where a mix of categorical and quantitative explanatory variables predict a quantitative response), variable transformations (e.g., logit transformations; polynomial model fitting), stepwise regression approaches, factors and blocking in randomized experiments, modeling fitting checking parameters, and so forth.

Following the general principles we have embraced in our version of the Stat 101 course (let the main focus drive all content, spiral over key concepts again and again, let conceptual framework and approaches take precedence over mathematical formulations, etc.), we have designed a preliminary version of a second course. In our preliminary version, we have developed chapters which immediately follow our Stat 101 (randomization-based) course chapters. In summary, the Stat 101 chapters move through three units: Unit #1 (Ch 1-4): The four pillars of statistical inference (Significance, Estimation, Generalizability and Causation), Unit #2 (Ch 5-7): Analyzing two groups (proportions, means, and paired data) and Unit #3 (Ch 8-11): Analyzing more general situations (ANOVA, Chi-square and Regression). In our preliminary version of Unit #4: Confounding and Variation—Two substantial hindrances to drawing conclusions from data, we have developed the following 7 Chapters.

Unit #4: Confounding and Variation---Two substantial hindrances to drawing conclusions from data (Second course material)

Ch 12: Stepping back and identifying the two major themes of statistical analysis (confounding and variation)- a review chapter which keys on the ideas of confounding and variation as they have been introduced and explored throughout Units #1, 2 and 3.
Ch 13: Blocked study designs to assess confounding and variability- a chapter which introduces blocking as an explicit way to address confounding (by limiting within group variability) for measurable confounding variables, instead of leaving their values “up to chance”; reviews paired designs for observational data as a corollary to experimental blocking
Ch 14: Analysis of confounding variables (simulation, ANOVA, and regression)- a chapter exploring simulation and theory-based approaches to comparing multiple groups in block designs; considers ANOVA as an approach to attribute variation to different variables; introduces indicator variables/regression model as an equivalent alternative to ANOVA, that also allows for analytic control of confounding variables in observational data.
Ch 15: Factorial designs to address variability and confounding- a chapter extending the idea of designing randomized experiments and observational studies to situations with multiple explanatory variables of interest
Ch 16: Analysis of factorial designs and observational studies with confounding variables- a chapter extending the ANOVA and regression approaches to experiments and observational studies with multiple explanatory variables of interest
Ch 17: Special topics in multiple regression (linear trend test, multi-category explanatory variables, step-wise regression etc.) – a chapter focusing on special applications and uses of the multiple regression model
Ch 18: Variable transformations in multiple regression models (response and explanatory variable transformations)- a chapter focusing on variable transformations and their utility in multiple regression models

The pedagogical focus of the second course continues our general approach to the first course, which uses guided exploration and

4. Describe preliminary results from implementation of such a curriculum

We developed and implemented such a curriculum in spring 2013 with 68 students in 3 sections of the class, at two institutions taught by two different instructors. Successful completion of the course as indicated by performance on a final exam, multiple class data design and analysis projects and a newly developed assessment test suggest that the curriculum has potential for preparing students well for the ideas of confounding and variation, while sacrificing little in terms of technical ability to design and analyze multivariable research studies and interpret multivariable research studies. We successfully implemented the curriculum with two different groups of students (1) a group of students who took a semester long randomization-based Stat 101 course and (2) a more mathematically mature group of students who took a half-semester, accelerated randomization-based stat 101 course (providing them the entirety of Units 1-4 in a single semester).

5. Open questions, next steps, and challenges to the field

Recent advances in understanding student learning in statistics and assessment, along with unprecedented changes in the methods used in statistical practice and a rapidly growing Stat 101 audience have all contributed to providing the statistics education community a grand opportunity to reshape Stat 101 and beyond. We have argued that most current Stat 201 courses focus on the concept of model, yet other options may exist. We argue that the concepts of confounding and variation are multivariable concepts that students should deepen their understanding of, and that models are a tool to provide that enhanced understanding. We have developed preliminary materials for such a course, and have observed positive student outcomes. However, many important questions remain.

At the forefront of these questions is whether confounding and variation are, in fact the concepts most important for students in post-Stat 101 courses. Running counter to argument that they are, is recognition that for purposes of description or prediction, confounding and variation may not be the most important concepts. This is an area for future discussion and debate among statistics educators. Furthermore, although we have developed an outline and preliminary materials for such a course and seen reasonable assessment results to date, given the preliminary state of the developed materials and that the second course itself is a relatively newcomer to the statistics education field, little is known about student performance and retention in such courses. A paramount need, in conjunction with community consensus about content in a second course in statistics is the development of valid and reliable assessment instruments and items. We have developed a preliminary assessment tool, but further reliability and validity testing is needed. Importantly, if the goal of the second course can be summarized by saying students should better understand confounding and variation, this presumes a basic level of understanding is obtained by students in the first course. What this understanding is and how to build on it well remain important questions worth answering. Our approach to the first course builds a stronger foundation of students' conceptual understanding which should enhance development of second course content. Finally, numerous implementation questions exist (e.g., randomized complete block design first or generalized block design? Connect the ideas of blocking and stratified sampling?).

6. Conclusions

Rapid growth in Stat 101 enrollments combined with a dramatic rise in the use of multivariable statistical approaches have presented the statistics education community with an unprecedented opportunity to think about the second course in statistics. Initial approaches have reduced the barriers to entry for such a course by developing curricula that do not rely on knowledge of calculus and linear algebra. However, in such courses statistical models become a conceptual focus. We argue that the motivation for such models, addressing confounding and variation through design and analysis strategies should be the focus, with models as an end themselves. We have outlined, developed and pilot-tested such a curriculum with promising initial results. Further discussion of conceptual goals for post-Stat 101 students are needed in order to drive further curricular innovation.

References

ARTIST (2005). *Assessment Resource Tools for Improving Statistical Thinking*. Retrieved from https://app.gen.umn.edu/artist

Cannon, A., Cobb, G.W., Hartlaub, B.A., Legler, J.M., Lock, R.H., Moore, T.L., Rossman, A.J. and Witmer, J. (2013) Stat 2: Building models for a world of data. WHFreeman.

Chance, B.L. and Rossman, A.J. (2013) Investigating statistical concepts, applications and methods. Second edition. http://www.rossmanchance.com/iscam2/index.html

Chihara, L.M. and Hesterberg, T.C. (2011) Mathematical statistics with resampling and R. Wiley.

Cobb, G.W. (1992), Heeding the Call for Change: Suggestions for curricular action (MAA Notes No. 22), chapter "Teaching Statistics" pages 3-43. The Mathematical Assocaition of America, Washington, DC, 1992.

Cobb, G.W. (2007). The introductory statistics course: a Ptolemaic curriculum? *Tech Innov in Stat Ed.* 1(1)

Cobb, G.W. (2011). Teaching statistics: some important tensions. *Chilean Journal of Statistics*, 2(1)

Guidelines for Assessment and Instruction in Statistics Education (GAISE) (2005). Retrieved from http://www.amstat.org/education/gaise

Garfield, J. (1995). How students learn statistics. *Intl Stati Rev*, 63: 25-34

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *Intl Stat Rev*, 75(3): 372–396

Kaplan, D. (2012) Statistical modeling: A fresh approach. http://www.mosaic-web.org/go/StatisticalModeling

Kuiper, S. and Sklar, J. (2013) Practicing Statistics: Guided investigations for the second course. Pearson.

Lock, R., Lock, P.F., Lock-Morgan, K., Lock, E.F. and Lock, D.F. (2013) Statistics: Unlocking the power of data. Wiley.

Lohr, S. (2011). "New Ways to Exploit Raw Data May Bring Surge of Innovation, a Study Says." *New York Times*, May 13, 2011. Retrieved at http://www.nytimes.com/2011/05/13/technology/13data.html

Nolan, D. and Temple Lang, D. (2007), "Dynamic, interactive documents for teaching statistical practice." *International Statistical Review*, 75(3): 295–321

Ramsey, F. and Schafer, D. (2013) The Statistical Sleuth: A course in methods of data analysis. Duxbury.

Switzer S. and Horton N. (2007). "What Your Doctor Should Know about Statistics (but Perhaps Doesn't)". *Chance*, 20(1): 17-21

Tabor J. and Franklin C (2011) Statistical Reasoning in Sports. WHFreeman.

Tintle, N.L, Chance, B., Cobb, G., Rossman, A., Roy, S., Swanson, T. and VanderStoep, J. (2013) Introduction to Statistical Investigations. Preliminary edition. http://math.hope.edu/isi

Tintle, N., VanderStoep J., Holmes V-L., Quisenberry B. & Swanson T. (2011). Development and assessment of a preliminary randomization-based introductory statistics curriculum. *J of Stat Ed*, 19(1)

Tintle, N., Topliff, K., VanderStoep J., Holmes V-L. & Swanson T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum *Stat Ed Res J*