Gateway Sciences Initiative Final Report

Grant Title: Statistics through Case Study

Project Summary:
The primary goal of the Statistics through Case Study Gateway Sciences Project was to implement a new paradigm in the teaching of introductory statistics, with the effect of increasing student mastery and retention of course material; fostering integration of statistics into students’ broader academic work; increasing student enthusiasm for statistics and appreciation of its widespread utility; and training teaching assistants and faculty in the implementation of a new approach to statistics pedagogy.

Our introductory statistics course is a service course taught to a wide variety of students, for many of whom it is a distributional requirement. Traditionally, the course was taught through large lectures in which the focus was primarily on the underlying theory. While this framework is effective for some students, it left a number of others unmotivated and unable to see the relevance of statistics to their own major disciplines, for example, or the necessity of statistical literacy in our larger world.

The new introductory statistics course 550.113, which we developed in this Gateway Sciences Initiative project, covers all of the topics from the traditional introductory statistics course 550.111, but does so through an active-learning approach, in which well-chosen case studies—born out of specific questions from psychology to epidemiology to political science—drive the development of statistical theory and methodology. The case studies have thematic threads that connect through the semester, and they integrate multiple segments of course material; as students work through them, they are forced to confront manipulation of real-world data, to complete significant reading of reference literature, to construct appropriate hypotheses, and to synthesize multiple concepts within statistics and from other disciplines. By design, several of these case studies cover research conducted by faculty in other departments at Johns Hopkins itself.

Indeed, that was another central goal of the project: to design and collect a database of suitable, often Hopkins-centered, case studies, resources, and materials for 550.111, the subsequent course 550.112, and a number of upper division courses as well. The modular structure of the case studies renders them appropriate for insertion into several different courses, and our evolving database of case studies has been fruitfully employed in classes
from the freshman to senior level. Finally, in keeping with our overarching goal not only of vertical integration of a new approach to statistics, but also department-wide reflection on and transformation of pedagogy, we conducted teaching and curriculum development workshops to train instructors and teaching assistants to effectively deploy new methods and stay informed about best practices.

**Major Impact and Continuity:** The new active-learning, case-study driven introductory statistics course (550.113) has been a significant success. In fact, the case studies and project modules have been commandeered for use by instructors of many of our other statistics courses, including higher-level calculus-based probability and statistics. The current departmental plan is to transform the year-long introductory statistics sequence in the mold of 550.113. Our curriculum development and teaching assistant training workshops now run annually, with a sustainable model of experienced faculty and graduate students involved in the training of new instructors. Finally, the evolving collection of case studies, refined as needed and easily incorporated into multiple courses, serves as a foundation for the implementation and endurance of our new paradigm.

**Participants:**

(a) *Faculty members:* Avanti Athreya, Donniell Fishkind, Bruno Jedynak, Daniel Naiman, and Fred Torcaso

(b) *Graduate and Undergraduate Teaching Fellows:* Sancar Adali, Deirdre Blanos, Florence Lau, Sherwin Lott, Zachary Lubberts, Matt Molisani, Daniel Sussmann, Steve Chestnut

**Activities Summary:** Work on the project began with a workshop in June of 2012 to strategize, collect resources, develop curriculum, and train instructors and teaching assistants. After extensive discussion and literature review of research on learning styles (complete with micro-teaching exercises for the workshop participants), as well as new classroom instructional techniques and their associated challenges, we assessed which research topics— with an emphasis on Johns Hopkins research programs—could be accessed and incorporated into our case studies. We gathered materials, conducted interviews with researchers whose work we targeted for inclusion in the case studies, and expended considerable effort writing and editing modules of case studies. Our aim in the case studies was to formulate well-
chosen “big” questions for the students to consider, find appropriate real data or design activities in which students collected data, and then push students to address the important central questions in each case study longitudinally through the semester as they acquired more statistical machinery. We developed appropriate classroom and homework exercises, and prepared assignments which familiarized students with state-of-the-art statistical computing. In fact, we held sessions in which teaching assistants were trained in the instruction of the R software package. We compiled all of these constituent parts into a comprehensive database of case studies and instructional material.

In addition to the workshop in June, we held a two-day workshop in August, the latter of which has since grown into a teaching assistant orientation and training that we repeat annually. The August workshop is focused on training instructors and teaching assistants, and includes discussion of lesson plans, learning styles, classroom management, implementation of active-learning approaches, and best practices in statistics pedagogy. The August workshop always concludes with microteaching presentations, which new teaching assistants frequently cite as “very useful.”

Incubated through these summer workshops, the Gateway Science Initiative course 550.113, Statistics through Case Study, debuted in the Fall of 2013 to a class of twenty-three students. It was a rich and rewarding course to teach, and student evaluations indicate that the students also regarded the course as stimulating and worthwhile. 550.113 covered the same theoretical material as 550.111, Statistical Analysis I; in addition, as part of the new course, students completed six case studies, drawn from psychology, epidemiology, economics, and political science. The case studies included manipulation of real data, readings from peer-reviewed literature, and synthesis of theoretical concepts. Students were introduced to the basics of the statistical programming language R and were able to use R for data analysis, random sampling, the construction of confidence intervals, and simple hypothesis tests.

The course opened with a case study on Stanley Milgram’s famous psychological experiment, a topic that animated student discourse. Students read Milgram’s original paper and manipulated his data; they discussed what they perceived as strengths and weaknesses of his experimental design and the relevance of his results to the present day. As an organic part of this process, students formulated questions about the “willingness” of certain individuals to inflict harm on others, and these questions led them directly to consider basic principles of conditional probability well in advance of any formal introduction to conditional probability itself. Students then read a Chance magazine article on the question of whether boys or
girls “run” in families, in the context of which they learned about basic discrete distributions; students used this knowledge to critique the authors’ methodology. Next, students considered the public health implications of disease testing, specifically how to interpret medical test results for a rare disease. The computations of posterior probabilities proved surprising—and illuminating—to many of them. To synthesize theoretical and practical perspectives on disease testing and public health, students read and discussed a 1995 paper in the American Journal of Epidemiology on estimating disease prevalence. Employment, particularly employment discrimination, was the focus of yet another case study, in which students used the hypergeometric distribution and Chebyshev’s inequality to investigate a claim of job discrimination in the US Postal Service, following a model developed in a 2000 paper in the Journal of Business and Economic Statistics. Since 2012 was an election year, students read Nate Silver’s popular “Five Thirty Eight” blog, and examined how to estimate the probability that a single vote might be decisive in an election, particularly in the context of a 2012 paper in *Economic Inquiry* which presents a mathematical model for the resolution of that very question. Our case studies also pushed students to learn the basics of computing with R, and to use R for simulation. More specifically, students performed repeated random sampling with R, determined the approximate distribution of the sample mean of several independent, identically distributed random variables, and interpreted the Central Limit Theorem. The ability to see, in real time, changes in the distribution of the sample mean also led several students to conclude that the sample size necessary for the Central Limit Theorem to be applicable depends on the underlying distribution itself. We emphasize that this insight is not one we typically expect of 111 students, and it arose seamlessly in the context of the activity in which they were engaged. Finally, students worked with data from Johns Hopkins University professor Justin Halberda’s lab on the idea of “number sense” in humans and they read a corresponding article in *Nature* by Halberda and his coauthors. In the investigation of the question of what kind of innate numerical “sense” humans possess, students used R to analyze laboratory data and to perform descriptive and inferential statistics.

We wish to stress that the amount and quality of the reading—some of it complex reference literature for specialists at the graduate or postdoctoral level—and the degree of interpretation, manipulation of data, and synthesis of methodology we demanded from the students in the course reflected a sea change from typical expectations at this level. Students found the course challenging, expressing that they “had all the work of the traditional
course, but more stuff on top of that.” As we detail in our discussion of student evaluations, many of them were motivated and enriched by these higher expectations, and were able to see unequivocally the relevance of statistics to questions of direct interest to them.

Further, based on student response to questions posed by teaching assistants, we suspect a curious selection bias: several students indicated to their teaching assistants that they enrolled in Statistics through Case Study precisely because they were hoping that the course would be less mathematical and less demanding than its traditional counterpart and that they were surprised by the workload and the rigor. We feel that this had an impact in the assessment of the course, a point we discuss below.

Assessment of the course: The first iteration of 550.113 was a vital step in the right direction. We still contend with challenges on the optimal course structure: initially, we aimed for one case study topic, article, or R lab per week, but students found this overwhelming. Enough students struggled with conditional probability and probability distributions that, after the first exam, we took a several-week break from case studies and focused almost exclusively on theoretical concepts and traditional homework exercises. After the second exam, we completed the political science and “number sense” case study described above.

As a part of our assessment plan, the final exam for the course, which was common to both 550.111 and 550.113, was uniformly graded. The results on the common final indicated that the 113 students outperformed the 111 students on nine of the eleven questions. While a Wilcoxon test did not indicate that this difference was significant, we nonetheless observed an improved conceptual grasp of the material on the part of the 113 students, illustrated partly by the quality of their responses to two of the questions we perceived as more conceptually challenging. Moreover, since we suspect a selection bias in favor of less mathematically inclined students, the improved performance of the students in 113 is a positive indication of the effectiveness of this approach. Since university policy prohibited us from randomly assigning students to one or the other course, we had to contend with a very small sample size in 113, which certainly impacted the results of our assessment.

Student feedback on course evaluations of 550.113 were positive. Students ranked the overall quality of the course highly, with a mean of over 4 (out of 5) in both sections. This is unusual—typical ratings of the traditional statistics class hover at around 3.5 out of 5. Students repeatedly pointed to the demonstration of statistical methodology through applications as an important strength of the course. One student wrote that “the best part
about this course was that it was designed to give some context to the abstract mathematical concepts that were being learned. The case studies and even many of the homework problems tried to ground the computation we were learning in concrete applications, which made it easier to remember. I liked the way [the instructor] started out the class by talking about why doing statistics is important from an epistemological standpoint, and tried to give us an idea of the overall significance of each technique that we learned.”

Not all responses were so positive, of course. Students cited the workload, exam difficulty, time allocated to different topics, lack of time to become proficient in R, and (as it happens) the attenuation of some of the case studies midway through the semester as points for improvement. On the whole, however, their enthusiasm for the case study approach, and its effectiveness in their retention of material, was clear.

We note that case studies, and the more active-learning approach overall, have now been implemented by multiple instructors across multiple courses: 550.111 and 112, but also 310 and 311 (Probability and Statistics for Engineering), 211 (Probability and Statistics for the Life Sciences) and 550.430 (Introduction to Statistics). Not only is the program sustainable, it is thriving, and has resulted in a shift of our approach to the teaching of applied mathematics and statistics in general. Further, our teaching workshops now run each year, with faculty and experienced graduate and undergraduate teaching assistants instrumental in the training of the new cohort of teaching assistants. We are also working on a current plan to conduct a one-week large lecture classroom “flip” for 550.112, to allow for lecture time to be devoted to a real-data ANOVA; the results of that will inform a larger plan to facilitate more active learning and the development of more online lecture material for future iterations of 111 and 112.

**Challenges and ongoing work:** Of course, significant challenges remain. As instructors, we continue to strive—and struggle, at times—to motivate students in a required course, especially one that imposes very real mathematical demands. It is also a challenge to introduce enough theoretical and computational machinery to make the resolution of real-world problems feasible. The nature of large lectures, with their attendant anonymity, and the physical barriers of typical large lecture halls both pose a nontrivial hurdle for us. We hope to find innovative solutions to some of these issues. Our future goals also include incorporating more cross-disciplinary case studies and including more guest speakers from different departments, as well as administering a course-wide initial diagnostic test for students in
order to accurately gauge improvements in statistical literacy and comprehension.

Due to IRB review, we will not be publishing the results of our project externally. We note, however, that the GSI project was a tremendous galvanizing force within the department to change how we approached the teaching of introductory statistics and how we approached our role more globally: not only as teachers and researchers, but as evangelists for the power of mathematical models, the crucial impact of proper experimental design, and the necessity for statistical fluency.