Engaging Students in the Practice of Statistics Through Undergraduate Research

Debra L. Hydorn

Department of Mathematics, University of Mary Washington, Fredericksburg, VA

ABSTRACT

As statisticians, we engage in a variety of activities, some of which are regularly integrated into our undergraduate courses. However, the individual courses that comprise a mathematics or statistics degree program might not provide students with experiences in the broader range of activities that define the practice of statistics. To remedy this situation, faculty can consider developing and mentoring undergraduate research projects. This article briefly discusses the skills that comprise statistical practice along with some course and program options for helping students to develop these skills. Then, types of undergraduate research projects in statistics are described to help faculty generate ideas for projects they could mentor. Several examples of each type of undergraduate research project are provided.

1. Introduction

Undergraduate programs for statistics and related majors often include opportunities for students to gain experience in the practice of statistics beyond their regular course work. Examples from institutions with undergraduate statistics programs include research courses (Yesilcay 2000), capstone courses (Spurrier 2001), consulting courses (Villagarcía 1998; Boomer, Rogness, and Jersky 2007), and consulting centers (Legler et al. 2012). Each of these examples provides students with opportunities to apply what they have learned in their statistics courses and to develop communication and other skills needed for success as a statistician. At issue, however, is how statistics faculty at institutions without a statistics major or minor or a consulting center can provide similar experiences for their students, to engage them in the practice of statistics. Some statistics faculty have developed opportunities for mathematics and general education students to engage in the practice of statistics. For example, one of the consulting courses described by Boomer, Rogness, and Jersky (2007) is open to students who are not majoring in mathematics but who have completed an introductory statistics course. A data analysis course (Schafer and Ramsey, 2003) can also provide statistics majors and non-majors alike with practice in statistics skills beyond an introductory course. However, for statistics faculty teaching in departments where consulting or data analysis courses are not available, mentoring undergraduate research projects can provide opportunities to expose students to the wider array of activities associated with the practice of statistics.

In this article, I will describe how undergraduate research can be used to engage mathematics majors in the practice of statistics after completing the first course in a probability and statistics sequence. Das (2013) describes a similar academic-year research program for mathematics majors with projects in graph theory, probability theory, actuarial mathematics, and statistics. My main goal for the undergraduate research projects that I mentor is to attract students to the statistics discipline, as advocated by Schafer and Ramsey (2003). I will provide a summary of the skills students need to be effective statistical practitioners, as well as recommendations for statistics faculty interested in mentoring undergraduate research projects and a summary of the overall benefits of undergraduate research. In addition, I will also provide a categorization of statistics undergraduate research projects gleaned from published research articles to serve as a basis for statistics faculty to generate their own undergraduate research projects in statistics. Although many authors have provided summaries of projects they have mentored (e.g., Villagarcía 1998; Yesilcay 2000; Delzell 2012), the categories of projects proposed here will provide interested faculty with a useful summary of the different kinds of undergraduate research projects they can consider.

2. The Practice of Statistics

In describing their research, consulting and other courses or programs, several authors have indicated what aspects of statistical practice their programs emphasize. Legler et al. (2012) provided examples of how their consulting center meets recommendations for training students in the practice of statistics that were set forth by the American Statistical Association’s Undergraduate Statistics Education Initiative.
Toward developing students’ statistical skills, Spurrier’s (2001) capstone course includes a series of eleven statistical methods-based experiences supplemented by nonstatistical modules to develop students’ communication and consulting skills. Because the projects differ each time a research project course is offered, the skills described by Yesilcay (2000) are project dependent. However, they typically include choosing an appropriate statistical model, data collection and analysis, and interpretation of results. Boomer, Rogness, and Jersky (2007) and Legler et al. (2012) described a similar reliance on the flexibility of statistical content needed for a successful consulting experience for their students. Das (2013) included some examples of specific mathematics skills and substantive area skills (e.g., background knowledge of actuarial mathematics) that were needed by students to complete their undergraduate research projects. While each of these experiences develops students’ skills in different ways, they share a common theme of working to solve a real problem. The components of statistical practice were further described by Pfannkuch and Wild (2000) in an article summarizing a series of interviews with professional statisticians on their working experiences. The PPDAC (Problem, Plan, Data, Analysis, Conclusions) framework used in this article to describe these statisticians’ experiences provides an effective structure for conducting undergraduate research in statistics.

While data analysis, modeling, and consulting are a big part of what statisticians do, one statistical activity that is not a part of the undergraduate research experiences described above is the development and/or evaluation of new or improved statistical methods. The undergraduate research projects that I have mentored often provide students with this experience. For example, one student I mentored explored the effect of modifying a regression dataset with the problem of multicollinearity by adding new data points and compared the results with other remedial methods for multicollinearity. Although this approach to dealing with multicollinearity may not be a practical option, the student learned about the effectiveness of remedial measures for multicollinearity and about the importance of designing an appropriate data collection protocol. Another student compared several methods for determining the number and type of parameters needed for fitting a time-series model to a collection of simultaneous datasets consisting of Twitter “tweets” from cities across the United States. Recently, two of my students created a modified form of Fleiss’s kappa for situations where raters made more than one rating on the same set of objects. They then evaluated the usefulness of their kappa measurement on a dataset with nine different rating criteria and compared the results to using Fleiss’s kappa on the nine ratings individually. Each of these research projects involved developing a new method or measure and then conducting simulations to evaluate their method or measure and to compare it to existing methods.

Similar to the program described by Legler et al. (2012), I try to provide research opportunities that appeal to and are accessible to mathematics majors. Undergraduate research projects that emphasize mathematical derivations and simulations, rather than data analysis and statistical applications alone, are accessible even to mathematics students with limited exposure to statistical methods and statistical thinking. The fact that the discipline of statistics combines a wide variety of activities and skills, from theoretical and mathematical results to applications and computational methods, is something that I try to demonstrate to my research students with each project. These projects can include some original work toward filling a gap in statistical knowledge, but that has not always been a requirement for my projects, contrary to the recommendations of Das (2013) and some definitions of undergraduate research (e.g., Roberts 2013). In addition, there is potential for undergraduate research projects in statistics to provide students with substantive content area skills, particularly through interdisciplinary collaborations. Examples include mathematical modeling (McMillan and Lickley 2008), biological systems (Diaz et al. 2009, and Friedman-Gerlicz 2009), environmental data (Carlson and Ecker 2002), physical properties (Senko 2010), and social structures (Egesdal et al. 2010). In each of these examples, substantial knowledge of the discipline underlying the data or system is required for students to place their results in the appropriate context.

3. Benefits of Undergraduate Research

Defined as a “High Impact Educational Practice” by the American Association of Colleges and Universities (http://www.aacu.org/leap/hip.cfm), undergraduate research is one form of active learning and much has been written on its benefits as used across many disciplines (see, e.g., Russell et al. 2007, Petrella and Jung 2008, and Lopatto, 2010). Within the mathematical sciences and statistics disciplines several of the authors cited in this article provide some discipline-specific and other benefits. Yesilcay (2000) provided a summary of benefits for students, from problem definition and model selection to developing students’ capacity for independent study, teamwork, and leadership skills, as well as potential employment opportunities for students. Das (2013) described the development of connections between students and the mathematics community resulting from participating in undergraduate research and the importance of academic year research projects for faculty who are not associated with a summer research program. Roberts (2013) included the additional benefit for faculty engaged in undergraduate research of providing a break from other academic responsibilities. For many faculty mentoring undergraduate research provides an opportunity to advance their research and/or work with statistical methods and tools that are not part of the classes they teach. Undergraduate research is also a great way for faculty to share their love of learning and their excitement for discovering something new. Similar to impacts described by Legler et al. (2012), most of the students I have worked with have gone on to complete an honors project based on their research with me or have asked me to mentor them in a project of their design. All but two have gone on to a graduate program in mathematics or statistics, with several earning a Ph.D. in statistics. Undergraduate research in statistics, like data analysis and consulting projects, has the potential...
to provide students with training in “the full range of skills necessary for successful application of statistics” and to demonstrate to students “what is exciting about the field of applied statistics” (Schafer and Ramsey 2003). If conducted as part of a research or consulting course, students may also use their undergraduate research experience to meet a major or minor requirement, depending on the requirements at their institution. My institution includes an experiential learning requirement as part of our general education program and many students use their undergraduate research experience to meet this requirement. While some disagreement exists among the authors cited here about the degree to which an undergraduate research project in statistics must involve original work to fill a gap in knowledge, all would agree with the “ultimate goal” of undergraduate research for students as stated by Roberts (2013), “to enrich their educational experience and to develop their interest in scholarly work.”

4. Undergraduate Research in Statistics

In addition to promoting the USEI guidelines, Legler et al. (2012) also emphasized the experiential learning aspects of undergraduate research and the importance of promoting interdisciplinary projects to prepare students for working as part of a team of scientists. They also provided a timeline for research activities, starting with literature review and data cleaning and ending with poster presentations and a research log. Delzell (2012) described a phase structure for conducting research with students (data acquisition, visualization, analyses, and communication). She also provided helpful hints, potential pitfalls, and alternative applications based on her experiences. The PPDAC framework used to describe statisticians’ experiences in Pfannkuch and Wild (2000) also provided an effective structure for conducting undergraduate research in statistics. Das (2013) recommended six steps for conducting undergraduate research: (1) discover a gap in knowledge, (2) literature search, (3) analysis of the problem (brainstorming), (4) develop method, (5) perform study, and (6) peer review.

Several authors have provided guidance on developing and mentoring undergraduate research, either in mathematics or in statistics. Roberts (2013) described issues with the associated problems of selecting good students and finding an appropriate problem. He provides a series of questions to consider when selecting students and approaches to developing research projects. Yesilcay (2000) discussed the preliminary work and activities needed to find projects and prepare students for conducting research. Project development begins during the previous academic year by making contact with government offices and other potential sources for data. Prior to the academic year research project, students may participate in a voluntary summer “apprenticeship in statistics” with the agency they will be working to become familiar with the agency and the research problem.

5. Types of Undergraduate Research Projects in Statistics

A review of undergraduate research projects in statistics from several sources revealed five different types of projects based on the types of activities students carried out. (Sources that were used include Yesilcay (2000), a list of Undergraduate Senior Statistics Abstracts from Robin Lock’s site at St. Lawrence University http://it.stlawu.edu/~lock/iussa/, the SIAM Undergraduate Research Online journal http://www.siam.org/students/siuro/, the American Journal of Undergraduate Research http://www.ajur.uni.edu/, The College of New Jersey Journal of Student Scholarship https://joss.pages.tcnj.edu/, and the Pi Mu Epsilon Journal http://www.pme-math.org/) The titles and abstracts of the projects available at these sources were reviewed for statistical content and focus. Articles identified through this process were accessed and reviewed to identify project goals and methods used. Projects were grouped according to similar student learning and project outcomes. Projects that exhibited characteristics of multiple project types were further reviewed to identify the main purpose or goal of the project. The five project types are described below, each with a brief description of two or more representative projects.

5.1. Data Analysis Projects

The primary goal for this type of project is to apply a statistical model or method, such as regression analysis, to existing data in order to answer a research question by producing summary statistics or providing parameter estimates. An example for an environmental studies project (Carlson and Eck 2002) students worked with a team of scientists and statisticians to assess changes in water quality in two lakes. A variety of water quality variables were examined (e.g., phosphorous, dissolved oxygen, and turbidity) using discriminant analysis and Analysis of Covariance (ANCOVA) to compare the two lakes and to determine if the lakes had changed from 1999 to 2000. The project described by Delzell (2012) where students analyzed African conflict and climate data is another good example of this kind of project. Two other examples include the use of statistical methods to test for compliance to Benjamin’s law (Pike 2008) and the use of quantile regression and time series to analyze temperature changes over time (Leider 2012). Finally, Cooper, Kirskey, and Diaz (2015) described the use of regression analysis to evaluate the use of an algebra diagnostic test as a predictor for success in an introductory statistics course.

5.2. Observational Research Projects

For these projects, students design a survey and data collection protocol that they carry out with the goal of answering a specific research question using data analysis methods. Alternatively, the data can be collected observationally following a specified convention. As an example, students designed and implemented a survey to investigate the relationships between work-related attitudes, health and coping, and family issues among male and female managers and nonmanagers in a large manufacturing organization (Apperson et al. 2002). In another example, Senko (2010) described an observational study to evaluate the needs of hard of hearing students and to determine best practices for meeting those needs. The student conducted site visits at three different elementary schools and then used grounded theory analytic approach (Glaser and Strauss 1967) to analyze her field notes and make recommendations of
important factors for determining academic placement for hard of hearing children.

5.3. Experimental Research Projects

These projects are similar to survey research projects but students design an experiment to collect data to analyze rather than using a survey or observational protocol. Hubers et al. (2003) described an experimental study to examine how participant motivation and experience during the experiment impacts participants’ compliance in completing a meaningless task. Using a factorial design the students studied how motivation, required vs. voluntary participation, and being treated as a data producer vs. a coinvestigator impacted participants’ completion of the task. In another example project, students compared an agent-based model with a self-exciting point process to model gang activity (Egesdal et al. 2010). The two models were compared using a histogram analysis and the Akaike information criterion. Additional examples include a study to examine the impact of fat content in diet on health problems in mice (Diaz et al. 2009) and a comparison of teaching scores between participants in a professional development program and a control group of teachers (Campanelli and Dougherty 2010). Experimental projects can also be developed for students with a mathematics or engineering background. For example, León-Cázares and Xoconostle-Luna (2014) described an experimental project to evaluate the accuracy of a model of an electric vehicle. The students compared simulated racing data with experimental results.

5.4. Statistical Methods Projects

These projects differ from those focused on data analysis in that students work on the development and evaluation of a new statistical method, or make comparisons between a new method and established, traditional methods of analysis. For example, students may derive properties of new methods mathematically or discover them through simulations and applications. In Grimmer (2005) the student proposed a model to explain voting behaviors and compared this model to ones found through a literature review. Additional examples include the development of error bounds for hypergeometric probabilities (Jalal 2001), a new visualization of Fisher’s iris dataset (Benson-Putnins et al. 2011), and a comparison of different methods for generating a phylogenetic tree (Leung 2012). In addition, several projects involving statistical methods that I have conducted were described above.

5.5. Probability Projects

The primary goal of these projects is to answer a question involving the probability of some event or to make predictions using a probability model. Students may derive a probability distribution or use a known probability distribution in a new setting, then produce estimates and predictions using the distribution or through simulations. Examples include a study of the misclassification rates of hypertension (Friedman-Gerlicz and Lilly 2009) and the use of different random number generators for Monte Carlo simulations in mathematical finance (Pita-Juarez and Melanson 2011). The consulting project described by Villagarcía (1998) also fits in this category. For this project, engineering students conducted simulations to model the probability that power stations will be down and then used their results to simulate the power supply. Two projects that I have mentored involved simulations to evaluate a probability. For the first project, the student ran simulations to evaluate the probability of a single item being among the top seven items in each of three lists of 30 ordered items. For the second project, the student investigated the probability that fewer than two women would be hired in seven independent job searches using Bernoulli random variables. In each case, the students ran the simulations and then derived the corresponding probability distribution.

6. Discussion

A review of the example projects provided here demonstrates the different ways that undergraduate research can contribute to the development of the skills students need for the practice of statistics. Each of these projects provides experience in one or more of the USEI skills listed in Section 2. For example, many of the projects described above involved data and modeling in other disciplines, for example, environmental science, psychology, education, and biology, which require the development of substantive area skills for analyzing data and interpreting the results. In addition to a variety of different statistical methods, some of the projects also required substantial mathematical skills, such as engineering and actuarial mathematics. Besides software used for data analysis, many of the projects also required computing skills to conduct simulations and develop data visualizations. Finally, nonmathematical skills, such as oral and written communication and working as part of a team, are a common component of the projects described above. This review also suggests the importance of establishing connections with faculty in other disciplines where statistical methods are an important part of students’ preparation, for generating project topics and creating interdisciplinary collaborations.

The variety of projects listed here also demonstrates the wide array of sources for project data and mathematical and other topics where statistical methods are appropriate. With some thought and preparation it should be possible to design an undergraduate research project in statistics to connect the needs and interests of faculty and students, and to promote student interest in statistics as a career or graduate school option. Interested faculty could start by considering which courses they teach that might lead to potential interesting projects as well as to students with interest in and sufficient background for those projects. They might also consider how to structure the project and what research activities students can successfully complete within the time allotted for the research project. If a consulting, capstone, or other regular course offering are not an option, faculty might investigate offering undergraduate research opportunities through a directed study or similar course format. An example course syllabus and two project abstracts are provided in the Appendix (available in the online supplementary files).

For statistics faculty teaching at institutions without a major or minor program in statistics, it can be difficult to
provide interested students with experience in all of the components of effective statistical practice through existing statistics courses. Undergraduate research is one way to fill gaps in students’ preparation for employment as a statistician or for graduate school in statistics. In addition to increasing students’ exposure to the practice of statistics, participation in undergraduate research has other benefits for both students and their faculty mentors. The goals of this article were to encourage statistics instructors to consider developing and mentoring undergraduate research articles and to provide a summary of the different kinds of possible projects.

References

Apperson, M., Schmidt, H., Moore, A., and Grunbert, L., (2002), "Women Managers and the Experience of Work-Family Conflict," American Journal of Undergraduate Research [online], 1. Available at http://www.ajur.uni.edu/v1no3/Apperson.pdf

Benson-Putnins, D., Bonfardin, M., Magnoni, M., and Martin, D. (2011), "Spectral Clustering and Visualization: A Novel Clustering of Fisher’s Iris Data Set," SIAM Undergraduate Research Online. 4. Available at https://www.siam.org/students/siuro/vol4/S01075.pdf

Boomer, K. B., Rogness, N., and Jersky, B. (2007), "Statistical Consulting Courses for Undergraduates: Fortune or Folly?" Journal of Statistics Education, 15.

Campanelli, J., and Dougherty, K. (2010) "The Effects of Sustained Professional Development on Middle School Mathematics Teachers in Urban Schools," TCNJ Journal of Student Scholarship [online], 12. Available at http://joss.pages.tcnj.edu/flex/2010/04/10-Dougherty-and-Campanelli.pdf

Carlson, E., and Ecker, M. D. (2002) "A Statistical Examination of Water Quality in Two Iowa Lakes," American Journal of Undergraduate Research [online], 1. Available at http://www.ajur.uni.edu/CarlsonEcker.pdf

Cooper, A., Kirksey, J., and Diaz, R. E. (2015) "Can an Algebraic Diagnostic Test be Used to Predict Final Grades in an Introductory Statistics Course?," American Journal of Undergraduate Research [online], 12. Available at http://www.ajuronline.org/uploads/Volume%202012/Issue_2/AJURVol12Issue2Jan2015pp43_56.pdf

Das, K. (2013), "From Inquiry-Based Learning to Student Research in an Undergraduate Mathematics Program," PRIMUS, 23, 829–836.

Delzell, D. A. P. (2012), "African Conflict and Climate Data for an Undergraduate Research Project," Journal of Statistics Education, 20. Available at http://www.amstat.org/publications/jse/v20n3/delzell.pdf

Diaz, P., Gillespie, M., Krueger, J., Perez, J., Radebaugh, A., Shearman, T., Vo, G., and Wheatley, C. (2009), "A Mathematical Model of the Immune System’s Role in Obesity-Related Chronic Inflammation," SIAM Undergraduate Research Online, 2. Available at https://www.siam.org/students/siuro/vol2issue2/S01032.pdf

Egesdal, M., Fatbauer, C., Louie, K., and Neuman, J. (2010), "Statistical and Stochastic Modeling of Gang Rivalries in Los Angeles," SIAM Undergraduate Research Online, 3. Available at https://www.siam.org/students/siuro/vol3/S01045.pdf

Friedman-Gerlicz, C., and Lilly, I. (2009), "Misclassification Rates in Hypertension Diagnosis due to Measurement Errors," SIAM Undergraduate Research Online, 2. Available at https://www.siam.org/students/siuro/vol2issue2/S01031.pdf

Glaser, B. G., and Strauss, A. L. (1967), The Discovery of Grounded Theory: Strategies for Qualitative Research. New York, NY: Aldine Transaction.

Grimmer, J., (2005), “Awareness, Partisanship and the Post-Convention Bounce: A Memory-Based Model of Post-Convention Presidential Candidate Evaluations: Part II – Empirical Results,” American Journal of Undergraduate Research [online], 3. Available at http://www.ajur.uni.edu/v3n3/Grimmer%20pp%202015-24%20rev.pdf

Hubers, K. J., Graf, E. R., and Lantinga, S. B. (2003), "The Influence of Meta-Experimental Factors on Compliance and Attitudes: Participant Motivation and Experimenter Demeanor," American Journal of Undergraduate Research [online] 2. Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.518.3777&rep=rep1&type=pdf

Jalal, A. I. M. (2001), "Error Bounds Involving Almost-Binomial Approximations of Hypergeometric Probabilities," Pi Mu Epsilon Journal, 11, 187–193.

Legler, J., Roback, P., Ziegler-Graham, K., Scott, J., Lane-Getaz, S., and Richey, M. (2012), "A Model for an Interdisciplinary Undergraduate Research Program," The American Statistician, 64, 59–69.

Leider, J. (2012) "A Quantile Regression Study of Climate Change in Chicago,1960–2010," SIAM Undergraduate Research Online, 5. Available at https://www.siam.org/students/siuro/vol5/S01174.pdf

León-Cázares, F. D., and Xoconostle-Luna, D. (2014), "A Student Approach to a Mathematical Simulation of a Racing Electric Vehicle," American Journal of Undergraduate Research [online], 12. Available at http://www.ajuronline.org/uploads/Volume%202012/Issue_1/AJURVol12Issue1Aug2014pp20to26.pdf

Leung, C. (2012), "Bayesian Models for Phylogenetic Trees," McGill Science Undergraduate Research Journal, 17. Available at http://msurj.mcgill.ca/vol7iss1/Leung2012.pdf

Lopatto, D. (2010), "Undergraduate Research as a High-Impact Student Experience," Peer Review, 12. Available at https://www.aacu.org/publications-research/periodicals/undergraduate-research/high-impact-student-experience

McMillan, J., and Lickley, M. J. (2008), "The Potential of Tidal Power From the Bay of Fundy," SIAM Undergraduate Research Online, 1. Available at https://www.siam.org/students/siuro/vol1issue1/S01006.pdf

Petrella, J. K., and Jung, A. P. (2008), "Undergraduate Research: Import, Accuracy, Benefits, and Challenges," International Journal of Exercise Science, 1, 91–95. Available at http://digitalcommons.kwu.edu/cgi/viewcontent.cgi?article=10366&context=ijes

Pfannkuch, M., and Wild, C. J. (2000), "Statistical Thinking and Statistical Practice: Themes Gleaned from Professional Statisticians," Statistical Science, 15, 132–152.

Pike, D. P. (2008), "Testing for the Benford Property," SIAM Undergraduate Research Online, 1. Available at https://www.siam.org/students/siuro/vol1issue1/S01009.pdf

Pita-Juarez, Y. H., and Melanson, S. (2011), "European Option Pricing Using a Combined Inverse Congruential Generator," SIAM Undergraduate Research Online, 4. Available at https://www.siam.org/students/siuro/vol4/S01077.pdf

Roberts, G. E. (2013), "Conducting Mathematical Research with Undergraduates," PRIMUS, 23, 785–797.

Russell, S. H., Hancock, M. P., and McCullough, J. (2007), "Benefits of Undergraduate Research," Science, 27. Available at http://science.sciencemag.org/content/316/5824/548.full

Schafer, D. W., and Ramsey, F. L. (2003), "Teaching the Craft of Data Analysis," Journal of Statistics Education [online], 11.

Senko, L. (2010), "Addressing the Unique Needs of Hard of Hearing Students in Diverse School Settings," TCNJ Journal of Student Scholarship [online] 12. Available at http://joss.pages.tcnj.edu/files/2012/04/2010-Senko.pdf

Spurrer, J. D. (2001), "A Capstone Course for Undergraduate Statistics Majors," Journal of Statistics Education [online], 9.

Villagarcia, T. (1998), "The Use of Consulting Work to Teach Statistics to Engineering Students," Journal of Statistics Education, 6.

Yesilcay, Y. (2000), “Research Project in Statistics: Implications of a Case Study for the Undergraduate Statistics Curriculum,” Journal of Statistics Education, [online], 8.