PASSION-DRIVEN STATISTICS: A COURSE-BASED UNDERGRADUATE RESEARCH EXPERIENCE (CURE)

Lisa Dierker, Jennifer Rose and Valerie Nazzaro
Wesleyan University, United States
ldierker@wesleyan.edu

Passion-Driven Statistics is a course-based undergraduate research experience (CURE) that has been used in statistics courses, research methods courses, mathematics, and data science courses across a wide variety of disciplines. Funded by the National Science Foundation, Passion-Driven Statistics empowers students to take the lead in the decision-making and execution of data-driven inquiry and the production of new knowledge (Dierker et al., 2012). Students learn the programming required to construct, manage, analyze, and interpret data. Liberal arts colleges, large state universities, regional universities, medical schools, community colleges, and high schools have all successfully implemented the curriculum (Dierker et al., 2018, Rose & Dierker, 2019.

PREVIOUS RESEARCH

Published research has shown that students enrolled in Passion-Driven Statistics are more likely to report increased confidence in working with data and increased interest in pursuing advanced coursework (Dierker et al, 2018). This data-driven CURE was also shown to promote further training in data and applied statistics. Using causal inference techniques to achieve matched comparisons across three different statistics courses, students originally enrolled in Passion-Driven Statistics were significantly more likely to take at least one additional undergraduate course focused on statistical concepts, applied data analysis, and/or use of statistical software compared to students taking either an activity-based psychology statistics course or a traditional lecture-based math statistics course (Nazzaro et al., 2020). Further, Passion-Driven Statistics students took a larger number of additional courses compared to students originally enrolled in either of the comparison courses.

Because of the focus on programming, we have also compared enrollment of the data-driven CURE to traditional introductory programming experiences, revealing higher rates of female and under-represented minority enrollment compared to both a general introductory programming course and an introductory course representing a gate to the computer science major (Copper & Dierker, 2017). Overall, when compared to both traditional introductory statistics and introductory programming courses, the data-driven CURE has been shown to attract students from a much wider range of academic backgrounds as measured by Math SAT scores (Cooper & Dierker, 2017; Dierker et al., 2015).

A MULTIDISCIPLINARY MODEL OF INQUIRY

The staggering amount of available data across industries, non-profit organizations, and governmental agencies, coupled with the expectation that progress be based on data-driven decisions, has created a need for data analysis and interpretation skills to become as ubiquitous as reading or writing. Yet, available training provides few models that create the necessary dialogue across disciplines (Cobb, 2015; Horton & Hardin, 2015).

The Passion-Driven Statistics curriculum takes advantage of students' natural curiosity and provides a common language for approaching questions across numerous disciplines. This is achieved by asking students to develop their own research question from several large data sets representing ecology, astronomy, public health, education, economics, medicine, etc. Our constantly evolving archive includes accessible data sets selected for a) completeness and clarity of their documentation; b) size (the larger the better); c) diversity of variables measured; and d) whether the data set adds to the variety of represented disciplines. We have found this inclusive approach to be successful in allowing students with diverse interests, skills, and educational goals to widely communicate across disciplines on a variety of analytic issues and to make the connections that allow for an exploration of questions within and between disciplines.

In: EM Jones (Ed.), *Fostering Learning of Statistics and Data Science*
Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),
July 2023, Toronto, Canada. https://iase-web.org

DATA-DRIVEN REASONING THROUGH PROGRAMMING

Students headed for the modern data-driven workforce need to be able to think and perform flexibly in the context of real-world data. Programming and the use of formal code or syntax represents a central skill that greatly expands one's capacity for not only managing and analyzing data, but for engaging in quantitative reasoning and data driven inquiry at the highest levels (Horton & Hardin, 2015; Nolan & Temple Lange, 2009). While in some cases students in higher education settings may be able to enroll in a course that covers the practical aspects of using a single programming platform or software package, what is missing is a translatable framework that provides students with flexible skills that they can use to explore data and formulate and test their own empirical questions in diverse contexts and with different tools.

To this end, Passion-Driven Statistics provides students with meaningful, real-world experiences with programming through work with SAS, R, Python, Stata, SPSS and other analytic platforms. From reading in data (file input/output), to managing variables (variable types, creation, and modification), to selecting data subsets (indexing and control structures), performing descriptive and inferential analyses (using functions with named arguments), and generating graphs (graphics output), students actively use code to write and execute programs. In contrast to the contemporary pedagogical approach of introducing a single program with "canned" exercises, our model uses project-based programming opportunities across multiple platforms as a critical building block for quantitative reasoning (Horton, 2015). Our carefully developed supporting materials (freely available through our website and through Schoology, Coursera, and other learning management platforms) translate data analysis steps across the major analytic tools used within the modern work force.

Pre and post course surveys administered across diverse secondary education settings showed that among students entering the course without confidence in their coding skills, over half showed increased confidence in writing code to analyze data. Further, the data-driven project-based model has been found to attract higher rates of women (60.7% vs. 22.7%) and under-represented students (24.0% vs. 14.0%) compared to enrollments in a traditional introductory programming course (Cooper & Dierker, 2017). Rather than producing individuals who can think about data systems issues from a software-specific or discipline-specific perspective, our project-based model supports students in moving flexibly between data analysis platforms, starting with proficiency with one and continuing to build new proficiencies as additional projects are completed.

REACHING DIVERSE EDUCATIONAL ENVIRONMENTS

With funding from the National Science Foundation, we have designed a curriculum based on project-based learning theory and demonstrated the portability and effectiveness of this model in terms of increased confidence, and improved attitudes and interest in data analysis within postsecondary setting and among URM high school students.

A selection of colleges and universities representing diverse geographical and educational settings who have successfully implemented the data-driven, project-based model are presented in Figure 1.

We believe that this course can benefit other schools not only through dissemination of our model and experiences, but also by making our resources widely available. We are happy to share our course materials with others and encourage faculty to consider integrating project-based course content into their curriculum (http://passiondrivenstatistics.com)

Figure 1: A Selection of U.S. Universities Implementing the Passion-Driven Statistics Curriculum.

ACKNOWLEDGMENTS

The development of Passion-Driven Statistics and the supporting instructional materials has been supported by grant #0942246, #1323084, and #1820766 from the National Science Foundation, Improving Undergraduate STEM Education (IUSE).

REFERENCES

Dierker, L., Kaparakis, E., Rose, J. & Selya, A. (2012). Strength in numbers: A multidisciplinary, project-based course in introductory statistics. *Journal of Effective Teaching*, 12(2): 4-14.

Dierker, L., Woods, K., Singer-Freeman, K., Germano, K., Cooper, J.L. & Rose, J. (2018). Evaluating Impact: A comparison of learning experiences and outcomes of students completing a traditional versus multidisciplinary, project-based introductory statistics course, *International Journal of Education, Training and Learning*, 2(1), 16-28.

Rose, J.S. & Dierker L. (2019). Adapting the Passion-Driven Statistics curriculum for transition to high school, *Teaching Statistics*; 1–7. https://doi.org/10.1111/ test.12209

Nazzaro, V., Rose, J., Dierker, L., (2020). A comparison of future course enrollment among students completing one of four different introductory statistics courses. *Statistics Education Research Journal*. 19(3), 6–17, http://iase-web.org/ Publications.php?p=SERJ

Cooper, J., & Dierker, L. (2017). Increasing exposure to programming: A comparison of demographic characteristics of students enrolled in introductory computer science programming courses vs. a multidisciplinary data analysis course. *International Research in Higher Education*, 2(1), 92-100. DOI: 10.5430/irhe.v2n1p92

Dierker, L., Cooper, J., Selya, A., Alexander, J., Rose, J. (2015). Evaluating access: A comparison of demographic and disciplinary characteristics of students enrolled in a traditional introductory statistics course vs. a multidisciplinary, project-based course, *Journal of Interdisciplinary Studies of Education*. 4(1), 22-37.

Cobb GW. (2015) Mere Renovation is Too Little Too Late: We need to rethink our undergraduate curriculum from the ground up. *The American Statistician*. 69(4):266-82. https://arxiv.org/pdf/1507.05346.pdf

Horton, N.J. (2015) Teaching the Next Generation of Statistics Students to "Think with Data": Special Issue on Statistics and the Undergraduate Curriculum. *American Statistician*. 69(4):259-

Nolan D & Temple Lang D. (2015) Approaches to Broadening the Statistics Curricula. In: Shelley MC, Yore LD, Hand B, editors. Quality Research in Literacy and Science Education. London, UK: Springer; 2009. 357-381. https://link.springer.com /content/pdf/10.1007/978-1-4020-8427-0.pdf#page=368, Journal of Problem-Based Learning;3(1):6-28.

Horton N.J. (2015) Challenges and Opportunities for Statistics and Statistical Education: Looking Back, Looking Forward. *American Statistician*. 69(2):138-45. https://arxiv.org/pdf/1503.02188.pdf