

Integrating Data Analysis and Statistics Across Disciplines

Anna Bargagliotti

Department of Mathematics, Loyola Marymount University, United States.

Abstract

To respond to this increased attention to statistics in society and work force, as well as to contribute to improved career preparation for students, it is imperative that we foster data literacy in our university student population. This paper discusses 13 learning outcomes that describe Advanced Data Literacy for university level students across different disciplines.

Keywords: *Data Science; Statistics; Undergraduate Education.*

1. Introduction

The demand for people educated in statistics and data science has grown tremendously over the past decade. Jobs related to statistics are expected to grow by about 27% between 2012 and 2022 according to the Bureau of Labor Statistics (Bureau of Labor Statistics, 2013). Key attributes to be successful in such jobs are good computing, analytic and statistical skills, good communication skills, ability to work with real data, ability in storytelling with data both verbally and visually, and the ability to work as a team (Davenport and Patil, 2012).

To respond to this increased attention to statistics in society and work force, as well as to contribute to improved career preparation for students, it is imperative that we foster data *proficiency* in our university population. Following the recommendations of the American Statistical Association put forth in the new Curriculum Guidelines for Undergraduate Programs in Statistics (ASA, 2014), this paper discusses how statistics and data analysis bridges many disciplines and how the different discipline approaches can be integrated. In this paper, we outline a series of common learning outcomes across disciplines for achieving data proficiency on a university campus.

2. Statistics and Data Science Education at the Universities Across Disciplines

Universities typically have several different statistics course offerings across campus. Because it is very common to have statistics courses housed in different disciplines (e.g., mathematics, computer science, psychology, economics), the American Statistical Association (ASA) and Mathematical Association of America (MAA) have offered guidelines for teaching introductory statistics targeted at non-statistics departments (ASA/MAA Joint Statement, 2014). Often times these courses overlap and yet their prerequisite structures do not allow a student to move from a statistics course offered in one department to a more advanced course offered by another department. Departments, often rightfully argue that the type of statistical techniques needed are discipline specific and thus necessitate the offering of a course within a specific discipline.

Although specific techniques do vary from discipline to discipline, certain basic themes of working with data should be present in all courses. Three important, fundamental, and particularly timely themes are that students need to (1) be employing technology, (2) be exploring real data sets, and (3) be practicing communicating statistical ideas and results. Moreover, in all disciplines, statistics should be guided and taught through the statistical investigative process of formulating a question, collecting appropriate data to answer that question, choosing the appropriate analysis technique to answer that question, and interpreting the results to answer the question (Franklin et al, 2007). The material commonly taught in introductory statistics courses often focuses on techniques, but such

methods are often “necessary but not sufficient” for modern data science (Hardin et al, 2015; Ridgeway, 2015).

Several important reports have stated the need for students to work with real data. The Committee on the Undergraduate Program in Mathematics Curriculum Guide 2015 (CUPMC, 2015) states “Working mathematicians often face quantitative problems to which analytic methods do not apply. Solutions often require data analysis, complex mathematical models, simulation, and tools from computational science.” This report recommends that all mathematical sciences major programs include concepts and methods from data analysis and computing. The Guidelines for Assessment and Instruction in Statistics Education (GAISE) college guidelines also included working with real data as one of the necessary six components of structuring a statistics course (ASA, 2016). In addition, the recommendations of the ASA on undergraduate programs in data science include Real Applications and Problem Solving as two of their Background and Guiding Principles. They state programs should “emphasize concepts and approaches for working with complex data and provide experiences in designing studies and analyzing real data (defined as data that have been collected to solve an authentic and relevant problem)” (ASA, 2014).

As data science has been described as an intersection of statistics with computer science, when considering undergraduate preparation, one must consider how the use of software interplays with statistics. Regardless of the discipline, technological fluency has become a must for success in the workforce. Therefore, university statistics and data science courses must incorporate heavy use of technology and computing.

Working with technology, working with real data, and communicating results provide the unifying themes for statistics and data science course offerings across campuses. If a model existed for what courses should look like across different departments that were centered on such themes, then the door may open for students seeking statistics and data science beyond what their departments offer.

2.1. Teaching and Learning of Statistics and Data Science at Universities

Much research has been dedicated toward uncovering best practices for undergraduate students’ learning of statistics. This research has largely targeted introductory statistics courses. A focus of the research has been to identify statistics topics and concepts that students have difficulty learning (Garfield & Ben-Zvi, 2008). Often students rely on “plug-and-chug” methods to solve problems without understanding the statistical ideas being discussed (Chervany et al., 1977; Stroup, 1984; delMas, 2004).

As noted by Horton & Hardin (2015), little research has targeted statistics learning beyond the introductory course. Given the the rapid growth of statistics and data science, it is important for research to “catch up” and inform best practices in developing, implementing,

and assessing programs. In addition, understanding how data science and computing interplay with statistics program offerings is currently an under-researched topic. Data science, as described by Baumer (2015), is “an emerging interdisciplinary field that combines elements of mathematics, statistics, computer science, and knowledge in a particular application domain for the purpose of extracting meaningful information from the increasingly sophisticated array of data available in many settings.” Some researchers have successfully implemented data science courses at their institutions (e.g., Hardin et al, 2015, Baumer, 2015) with similar student learning goals.

2.2. Statistics and Data Science at Loyola Marymount University

The National Science Foundation (NSF)-funded project (NSF Grant No. 1712296), *Undergraduate Data Pathways* (UDaP), focuses on understanding differences and similarities of statistics and data analysis course offerings across different disciplines. Using Loyola Marymount University (LMU), a mid-sized comprehensive university in Los Angeles, California that offers 53 major and 57 minor undergraduate programs; 31 master’s degrees; one doctoral degree; and 15 credential programs as a case study, the project defined a set of learning outcomes that integrated the statistical goals put forth by several different disciplines.

While LMU does not have a department dedicated to statistics or data science, the Department of Mathematics, Department of Biology, Department of Engineering, Department of Economics, Department of Political Science, Department of Psychology, Department of Sociology, the School of Business, and the School of Education offer courses related to statistics and data analysis. Several of the introductory courses offered in these departments overlap in content thus creating an appearance of duplication and potential wasted resources across campus. Furthermore similar courses in one department are often not accepted as prerequisites for higher-level statistics offered in another department, thus blocking the pipeline for students to advance their statistical knowledge. Because of LMU’s breadth of programs, its size and relatively typical situation regarding statistical course offerings, the findings for LMU can provide a model for other universities wishing to unify statistics offerings across disciplines around working with real data, technology, and emphasizing communication.

3. Processes

Five steps have been undertaken to research the differences and commonalities of statistics and data analysis across disciplines. As a first step, a faculty working group with representation from mathematics, economics, biology, psychology, sociology, business, and statistics was formed. The working group was centered around understanding the processes and support needed to implement the themes of communication, technology, and real data

in statistics courses across the disciplines. Four meetings per semester have consistently been scheduled over the course of the last two academic years. The purpose of the working group discussions is to gather qualitative data on how the different disciplines articulated the importance of statistics and to reach common ground to understand what all of the disciplines had in common.

The second step in the process was to develop and administer a 15 minute survey to the working group. The survey asks about the software platforms they used, data sources they used, the types of class assignments they gave (e.g., statistics investigations in the form of projects, problem sets), and the types of activities they used in the classroom (e.g., students using computers in a lab setting, group work). The survey included questions from the NSF-funded Statistics Teaching Inventory (STI) surveys developed by Zieffler et al. (Zieffler, 2012) focusing on their teaching practice, assessment practice, teaching beliefs and assessment beliefs.

A third step was to review any position statements, policy documents, or curriculum guidelines written by professional organizations regarding data proficiency to understand whether there was common ground between the disciplines.

A fourth step included carrying out a survey to the community, both academic and non-academic, to garner their thoughts on the necessary learning outcomes for statistics and data analysis at the university level.

Lastly, the culminating step of the work was to develop a set of learning outcomes for teaching and learning data-related courses based on the findings of the prior steps. Thirteen such learning outcomes were agreed upon.

4. Findings

The Table 1 presents the final 13 learning outcomes that were established as important for students to meet at the university level. Students meeting these learning outcomes are deemed to be *Advanced Data Literate*.

The learning outcomes span both content and process. The important themes of using real data, communication with data, and technology are well-represented within the learning outcomes as well. These outcomes are meant to be broad and cross-disciplinary so they can serve as benchmarks across all disciplines offering statistics and data analysis courses. These learning outcomes stemmed from the extensive discussions within the working group as well as the review of the policy documents and the community survey.

Table 1. Learning Outcomes and Descriptions

Learning Outcomes	Descriptions
1	Students formulate and/or address questions about univariate data, collect/consider univariate data, analyze univariate data, and interpret results
2	Students understand, calculate and interpret descriptive measures for quantitative and/or categorical variables to describe characteristics of the data
3	Students create and interpret basic data visualizations for quantitative and categorical variables
4	Students understand, carry out, and interpret basic inferential statistical procedures for one or two samples
5	Students understand, carry out, and interpret results from estimating statistical models for bivariate data (e.g., linear regression, interpolation, extrapolation, predictive inference)
6	Student carry out and communicate results from extensive data-driven project(s) that is related to a real-life problem (extensive means that a single project takes more than two weeks to complete or a series of projects take more than two weeks to complete and are worth at least 25% of the final grade)
7	Students communicate their analyses and the interpretations of their results in a manner that is appropriate to their discipline in the context of the data (e.g., communication could be emphasized with presentations, oral explanations of results, oral/written answers for in-class work, written explanation of results)
8	Students understand the implications of study design, can select appropriate statistical methods for data analysis, and can explain limitations of their analyses and interpretations
9	Students become critical consumers of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted
10	Students formulate and/or address questions about multivariate data, collect/consider multivariate data, analyze multivariate data, and interpret results
11	Students use current statistical software or statistical packages that are appropriate to the discipline and context beyond basic Excel or a calculator
12	Students write a program (using a programming language) to analyze data or extract information from the data
13	Students study at least one type of advanced data-analytic methods such as (not limited to): generalized linear models, Bayesian analysis, advanced probability theory and stochastic processes, non-linear models, machine learning, advanced study-design, big data analysis, econometrics, or statistical computing

Policy documents. Six curriculum guidelines were found from various professional organizations that specifically discuss students’ necessary data literacy. Each of these documents was reviewed to cross-check whether the learning outcomes were explicitly mentioned. Five different disciplines are represented in the policy documents; Mathematics, Statistics, Psychology, Economics, and Sociology. The Table 2 presents the results.

Table 2. Professional Organization Alignment with Learning Outcomes

Learning Outcomes	1	2	3	4	5	6	7	8	9	10	11	12	13
American Statistical Association (ASA)	X	X	X	X	X		X	X	X	X	X	X	X
Mathematical Association of America (MAA)	X	X	X	X	X		X	X	X	X	X	X	X
Guidelines for Assessment in Statistics Education (GAISE)	X	X	X	X	X		X	X	X	X	X		
American Psychological Association (APA)	X	X	X	X			X		X	X	X		
American Economics Association (AEA)	X	X	X	X	X		X		X	X	X	X	X
American Sociology Association (ASA)	X	X	X	X		X	X		X	X	X		

Table 2 illustrates that eight of the learning outcomes were discussed in all of the policy documents. The remaining five learning outcomes were supported by three or more of the documents. Only in one case, the extensive project, was the learning outcome not well-supported. In this case, the documents did not discuss the need for a project, however, they also did not suggest that an extensive project was a bad idea. A project offers students a way to connect ideas in a coherent manner and further work on their communication. It also

provides an opportunity for students to work extensively with real data as well as technology to carry out the analyses.

To further validate the learning outcomes, a community survey was administered online. The online survey was sent out by members of the working group to their connections, was sent out of list serves for several disciplines, and was posted on a few forums. A total of 367 people opened the survey and 287 people completed the survey. The following pie graph show the distributions of backgrounds of people who completed the survey.

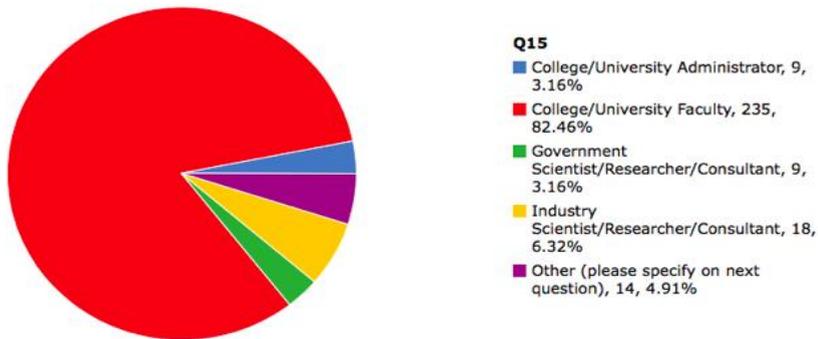


Figure 1. Distribution of backgrounds.

The next table shows the percentage of survey respondents that agreed, were neutral, or disagreed with the statement that the learning outcome was an important skill that a university student must acquire.

Of the 13 learning outcomes, four of them had 90% or above of the 287 people in agreement that they are important skills that a university student must acquire. Another five learning outcomes had a large majority of respondents state that they agreed or were neutral. Only three learning outcomes had large disagreements with the statements.

As a result of the findings, the 13 learning outcomes were deemed to describe an Advanced Data Literate student. A student meeting less of these outcomes could be labelled at lower levels of data literacy.

5. Conclusions and Future Research

The main goal of this paper is to outline 13 learning outcomes for statistics and data literacy at the university level. As society pushed towards being more data-driven, it is important to understand and characterize what education should be doing as a response. The common learning outcomes in statistics and data science across disciplines focused around three important themes in data science: working with real data, communication, and technology.

This project is ongoing and the findings are expanding. This paper offers the important initial step in finding common ground across disciplines. The creation of a working group of “change agents” on a university campus that have interest in furthering data proficiency in students has been an invaluable asset to the project. To be effective, these change agents must come from different disciplines. The formation of a working group of invested change agents is no easy task. For the formation of such a group at LMU, we sought the help of the Associate Dean for Undergraduate Studies who urged faculty that had investment in statistics to join the group. In addition, members of the research team personally reached out to faculty in other departments to encourage them to join the working group. In total, a working group of 10 faculty was set up.

Future work for the project includes reviewing enrollment data and all of the courses offered at LMU. The goal is to contribute a comprehensive picture of what is happening at the university level surrounding statistics and data science.

References

- American Statistical Association, Undergraduate Guidelines Workgroup. (2014). *Curriculum Guidelines for Undergraduate Programs in Statistical Science*, retrieved from <http://www.amstat.org/education/curriculumguidelines.cfm>.
- American Statistical Association. (April, 2014). *Statistical Research and Training under the BRAIN Initiative*, retrieved from http://www.amstat.org/policy/pdfs/StatisticsBRAIN_April2014.pdf.
- AMSTAT News, (2015). Statistics: Fastest-Growing Undergraduate STEM Degree. AMSTATNEWS, March Issue.
- ASA/MAA Joint Committee on Undergraduate Statistics. (2014). *Qualifications for Teaching an Introductory Statistics Course*, American Statistical Association, Mathematical Association of America, retrieved from www.amstat.org/education/pdfs/TeachingIntroStats-Qualifications.pdf.
- Baumer, B. (2015). A Data Science Course for Undergraduates: Thinking with Data, retrieved from <http://arxiv.org/pdf/1503.05570.pdf>
- Bureau of Labor Statistics. (2013). *Employment by Detailed Occupation, 2012 and Projected 2022 Table*, retrieved from www.bls.gov/emp/ep_table_102.htm.
- Cai, J. (2000). Understanding and representing the arithmetic averaging algorithm: An analysis and comparison of U.S. and Chinese students' responses. *International Journal of Mathematical Education in Science and Technology*, 31, 839–855.
- Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about Sampling Distributions, In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 295–323). Dordrecht: Kluwer.
- Chance, B. & Rossman, A. (2015). *Investigating Statistical Concepts, Applications, and Methods*. Thomson Brooks Cole, Third Edition.

- Chervany, N. L., Collier, R. D., Fienberg, S., & Johnson, P. (1977). A framework for the development of measurement instruments for evaluating the introductory statistics course. *The American Statistician*, 31(1), 17–23.
- Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly*, 104, 801–823.
- [Committee](#) on the Undergraduate Program in Mathematics (CUPM). (2004). *Undergraduate Programs and Courses in the Mathematical Sciences: CUPM Curriculum Guide 2004*. Mathematical Association of America. Available at www.maa.org/cupm/.
- delMas, R. C. (2004). A comparison of mathematical and statistical reasoning. In J. Garfield & D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 79–95). Dordrecht, the Netherlands: Kluwer Academic.
- delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28–58.
- Davenport, T., & Patil, D.J. (2012). Data Scientist: The Sexiest Job of the 21st Century. *Harvard Business Review*, 90(10), 70–76.
- Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, 28, 96–105.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K–12 curriculum framework*. Alexandria, VA: American Statistical Association. (Also available at www.amstat.org.)
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Berlin, Germany: Springer.
- Green, D. R. (1983). A survey of probability concepts in 3000 pupils aged 11-16 years. In D. R. Grey, P. Holmes, V. Barnett, & G. M. Constable (Eds.), *Proceedings of the First International Conference on Teaching Statistics*. Sheffield, England: Teaching Statistics Trust.
- Groth, R. E. (2002). Characterizing secondary students' understanding of measures of central tendency and variation. In D. S. Mewborn, P. Sztajn, D. Y. White, H. G. Wiegel, R. L. Bryant, & K. Nooney (Eds.), *Proceedings of the twenty-fourth annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education: Volume 1* (pp. 247–257). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.
- Groth, R. (2003). High school students' levels of thinking in regard to statistical study design. *Mathematics Education Research Journal*, 15(3), 252–269.
- Hardin, J., Hoerl, R., Norton, N., & Nolan, D. (2015). Data Science in Statistics Curricula: Preparing Students to “Think with Data,” working paper, accessed at <http://arxiv.org/ftp/arxiv/papers/1410/1410.3127.pdf>
- Hesterberg, T. C. (1998). Simulation and Bootstrapping for Teaching Statistics. *American Statistical Association Proceedings of the Section on Statistical Education*, 44–52. Available online at www.timhesterberg.net/articles.

- Hodgson, T. & Burke, M. (2000). On Simulation and the Teaching of Statistics, *Teaching Statistics*, 22 (3), 91-96.
- Holdren, J. P. and Lander, E. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. https://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf.
- Horton, N. & Hardin, J. (2015) Teaching the Next Generation of Statistics Students to “think with Data”: Special Issue on Statistics and the Undergraduate Curriculum. *The American Statistician*, forthcoming issue November.
- Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: Responding to classroom realities. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 909–955). Charlotte, NC: Information Age.
- Koehler, M. H. (2006). Using Graphing Calculator Simulations in Teaching Statistics. In G. F. Burrill & P. Elliott, (Eds.), *Thinking and Reasoning with Data and Chance, Sixty-eighth Yearbook* (pp. 257-272). Reston VA: National Council of Teachers of Mathematics.
- Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, 3(1).
- Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, 33(4), 259–289.
- Konold, C., Pollatsek, A., Well, A. & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE Roundtable Conference* (pp. 151–167). Voorburg, the Netherlands: International Statistical Institute.
- Lock, R., Lock, P., Lock Morgan, K., Lock, E., & Lock, D. (2012). *Unlocking the power of data*. Wiley.
- LMU Mathematics Department Careers Committee. (2014). Survey of Online Job Posting in California, Internal Report.
- Meletiou-Mavrotheris, M. (2003). Technological Tools in the Introductory Statistics Classroom: Effects on Student Understanding of Inferential Statistics, *International Journal of Computers for Mathematical Learning*, 8, 265-297.
- Mills, J. D. (2002). Using Computer Simulation Methods to Teach Statistics: A Review of the Literature, *Journal of Statistics Education* [online], 10(1). Available at www.amstat.org/publications/jse/jse_archive.htm#2002.
- Nolan, D. & Temple Lang, D. (2010). Computing in the statistics curriculum. *The American Statistician*, 53: 370-375.
- Nolan, D. & Temple Lang, D. (2015). Data Science in R: A case studies approval to computational reason and problem solving. *The American Statistician*, forthcoming November issue.

- Pfaff, T. & Weinberg, A. (2009). Do Hands-On Activities Increase Student Understanding? A Case Study,” *Journal of Statistics Education* [online], 17(3). Available at www.amstat.org/publications/jse/jse_archive.htm#2009.
- Piaget, J., & Inhelder, B. (1975). *The origin of the idea of chance in children*. London, England: Routledge & Kegan Paul.
- Ridgeway, J. (2015). Implications of the Data Revolution for Statistics Education, *International Statistical Review*, doi:10.1111/insr.12110
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). Charlotte, NC: Information Age.
- Shaughnessy, J. M., Ciancetta, M., & Canada, D. (2004). Types of student reasoning on sampling tasks. In M. Johnsen Hoines & A. Berit Fuglestead (Eds.), *Proceedings of the 28th meeting of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 177–184). Bergen, Norway: Bergen University College Press.
- Stroup, D. F. (1984). The statistician and the pedagogical monster: Characteristics of effective instructors of large statistics classes. *Proceedings of the Section on Statistical Education* (pp. 1–7). Washington, DC: American Statistical Association.
- Tintle, N., Chance, B., Cobb, G., Rossman, A., Roy, S., Swanson, T., & VanderStoep, J. (2014). *Introduction to Statistical Investigations*, Preliminary Edition. Wiley.
- Watkins, A., Bargagliotti, A. E., & Franklin, C. (2014). Simulation of the sampling distribution of the mean can mislead. *Journal of Statistics Education*.
- Wilder, E. I. (2010). A Qualitative Assessment of Efforts to Integrate Data Analysis throughout the Sociology Curriculum Feedback from Students, Faculty, and Alumni. *Teaching Sociology*, 38(3), 226-246.
- Zieffler, A., Park, J., Garfield, J., delMas, & R., Bjornsdottir, A. (2012). The statistics teaching inventory: A survey on statistics classroom practices and beliefs. *Journal of Statistics Education*, 2(1).
- Zorn, P., Bailer, J., Braddy, L., Carpenter, J., Jaco, W., and Turner, P. (2014). The INGenIOuS project: Mathematics, statistics, and preparing the 21st century. <http://www.maa.org/sites/default/files/pdf/ingenious/INGenIOuS-report.pdf>.