OVERVIEW OF AN EDUCATIONAL MODEL OF ELEMENTARY STATISTICS FOR THE DATA SCIENCE PROGRAM OF SHIGA UNIVERSITY

Shizue Izumi and Michinori Hatayama
Department of Data Science, Shiga University, Japan
Disaster Prevention Research Institute, Kyoto University, Japan
shizue-izumi@biwako.shiga-u.ac.jp

We propose a design for an elementary statistics course incorporating a massive open online course (MOOC), real-data-oriented project-based learning (PBL), and artificial intelligence (AI). In April 2017, Shiga University established the first department of data science in Japan. We intend to teach data science by enabling students to accumulate experience in solving real-world problems using various data and creating value from the data. The course contents include development of critical thinking, communication, and teamwork. Whereas students utilize MOOC to acquire knowledge outside the classroom, they acquire skills in communication and teamwork inside the classroom through PBL practice by examining data. We hope that our proposed model will help provide the students with a balanced introduction to statistical concepts, methods, and theory.

BACKGROUND

In April 2017, Shiga University established the first department of data science in Japan. Its curriculum consists of a series of courses in statistics and information science, which are approved by the Minister of Education, Culture, Sports, Science and Technology, Japan (Shiga University, 2017a). In 2017, 110 new students (81 male and 29 female students) enrolled in the program, for which the annual capacity is 100 students. Among the 110 students, approximately 60% chose Mathematics III (mathematics for science major), whereas the remainder chose Mathematics IIb (mathematics for liberal arts major) at their high schools.

We intend to develop data scientists having scientific skills and a liberal arts mindset According to the curriculum policies of the data science program (Shiga University, 2017b), students acquire expert knowledge in statistics and information science and are trained as data scientists specializing in data utilization. Our curriculum encourages students to increase their interest and willingness to work in the data science field, understand the importance of having a data science background, and accumulate experience in creating value from real big data through project-based learning (PBL) practices every four years (Takemura, et al., 2018).

Creating value in data science refers to how we attempt to solve a problem by defining it in various fields and taking advantage of analysis results to enhance decision-making in business and daily life (Maruyama, et al., 2015). In PBL practices, students consider each phase in the Problem-Plan-Data-Analysis-Conclusion (PPDAC) cycle and conduct a cross-evaluation of the group presentation using a checklist. Creating value implies developing the power to live, which is a goal of the current course of study set by the Ministry of Education, Culture, Sports, Science and Technology.

We propose a design of an elementary statistics course for the data science program offered by Shiga University. This is an undergraduate freshman program that aims to understand data analysis against a background of elementary statistics. We intend to teach data science by enabling students to accumulate extensive experience in solving real-world problems using various data and creating value from the data. The course contents encompass development of critical thinking, communication, and teamwork skills in students.

PREVIOUS EFFORTS

Pilot studies examined the feasibility of the proposed educational model in other national university in 2015 and 2016 (Izumi, 2017). Such studies were undertaken in some basic statistical courses among third-year students. WebClass, which is one of the most popular learning management systems (LMSs) in Japan, is used to set up an electric bulletin board, post video lectures, distribute course materials, submit a report and presentation, set up an electric portfolio, and conduct a cross-evaluation of the group presentation. One instructor and three teaching

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

assistants conduct these courses. In PBL practices, six students form a group, an instructor evaluates the submissions (reports and group presentations), and students participate in a question-and-answer session of the group presentation. A rubric is used to evaluate the submissions, and the results are instantly fed back to students through an electric portfolio to encourage them to reflect on what they have learned.

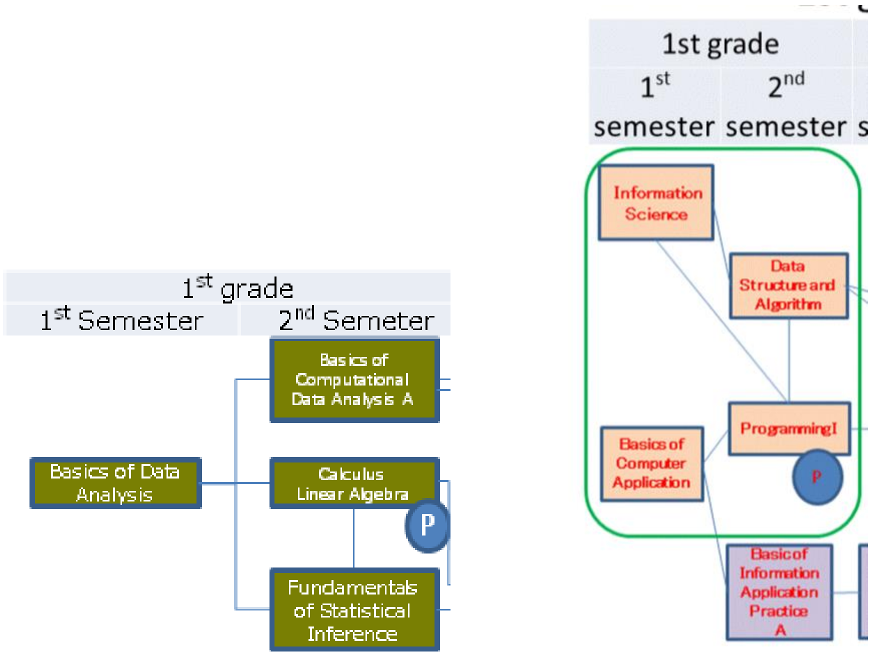

Figure 1. Courses on data analysis (left figure) and information (right figure) in the first grade's curriculum tree of data science program in Shiga University (Shiga University, 2017c)

COURSE DESIGN

Figure 1 shows the first grade's curriculum tree of data science program in Shiga University. Here, the term first grade refers to the first year of college. As part of courses on data analysis, freshmen learn real data handling and the basics of statistical thinking. In the course titled "Basics of Data Analysis," which is part of the first semester, freshmen learn about different types of data (discrete and continuous) and how to view a statistical graph. In addition, they learn cross-tabulation and descriptive statistics such as mean and correlation. Further, they briefly study linear regression and time-series data. In the course titled "Fundamentals of Statistical Analysis," which is part of the second semester, freshmen learn the concepts of population and sample, which are key to the development of statistical thinking. Further, to understand statistical estimation and hypothesis testing, students learn about the random variable and its distribution. In the second semester, they attend the calculus and linear algebra courses to build a sound mathematical knowledge foundation, which is important in understanding statistical theories.

These courses use an extensive range of active learning techniques. In both "Basics of Data Analysis" and "Fundamentals of Statistical Analysis," a massive open online course (MOOC) is used outside the classroom. On the other hand, during the second semester, the course titled "Fundamentals of Statistical Analysis" comprise face-to-face classroom activities, including a guest lecture, problem exercises, and PBL practices. One of the themes of the guest lecture in 2017 was data management in insurance mathematics. As part of PBL practices, five students form a group and accumulate experience in solving a problem using the PPDAC cycle. Our PPDAC cycle includes data cleansing in the Data phase and (written, oral, and visual) communication of results in the Analysis phase. In hybrid (blended)-style active learning, students express various statistical concepts using their own words (Izumi, et al., 2016). This type of courses enables more flexible learning than traditional lectures since the former involves lesser limitation of lecture time than the latter.

Further, the course titled “Basics of Computational Data Analysis A” teaches how to analyze data using the tools and functions available in Microsoft Excel. The course contents include computation of descriptive statistics, such as average, statistical estimation, hypothesis tests, linear regression analysis, and analysis of variance. In the first year of college, Microsoft Excel is chosen as the primary tool of data analysis for students.

The quality assurance of statistical education is performed by the Ministry of Education, Culture, Sports, Science and Technology and an external advisory board involving our department. In their fourth semester, all students take a grade 2 examination to qualify for the Japan Statistical Society Certificate.

Teaching facilities

The Department of Data Science encourages students to use laptop computers in the classroom and supports the software campus license agreement. Our computer environment includes security software to protect against computer viruses; Microsoft Office to prepare documents, tables, and presentations; Business Intelligent tools like Tableau to summarize data; and IBM Watson for cognitive computers. In addition, we prepare statistical data analysis software such as R, SPSS, and JMP/SAS, which are taught in advanced courses following the first year of college. The first-semester courses “Basics of Computer application” and “Theory of Data Science A” enable computer practice of Microsoft Office and offer a hands-on seminar on IBM Watson, whereas the second-semester courses “Programming I” and “Programming I Practice” teach basic skills in designing and writing computer programs using Python. Whereas a student uses AI tools with IBM Watson to solve a question using some example data sets from textbooks, a teacher uses these tools to analyze classroom data obtained from students and measure academic achievements. Web applets (e.g., Rossman / Chance Applet Collection, http://www.rossmanchance.com/applets/) are used to explain statistical concepts by utilizing Google’s auto translation of English to Japanese. Such facilities are nearly equivalent to those in American universities (Izumi and Gould, 2010).

Classes are maintained using an LMS based on Moodle, which is known as the Shiga University Learning Management System (SULMS). Students download course materials, answer surveys, submit reports, and join discussions in a forum using a web browser.

Japan Massive Open Online Courses

In 2013, Japan Massive Open Online Courses (JMOOC) were established with the aim of spreading the Japanese version of MOOC. The platform approved by JMOOC includes gacco by DOCOMO gacco; OpenLearning Japan by Net Learning, Inc.; and OUJ MOOC by the Open University of Japan (Sakurai, et al., 2015; Izumi, et al., 2016; Iyoku, 2017).

The course titled “Statistics I: Basics of Data Analysis” in JMOOC includes the chapters 1) Invitation to Statistics, 2) Statistical Graphs and Qualitative Data, 3) Quantitative Data, 4) Correlation and Time Series Data, and 5) Use of Public Statistics. Chapters 1 through 4 provide 10 quizzes each, while Chapter 5 provides 30 quizzes. Students having at least 42 (= $(10 \times 4 + 30) \times 0.6$) correct answers pass and receive a certificate (Sakurai, et al., 2015). This course has been created with the cooperation of the Japan Statistical Society (Takemura, et al., 2016).

The course “Statistics II: Methods for Statistical Inference” in JMOOC includes the chapters 1) Ideas of Statistical Inference and Probability, 2) Statistical Estimation, 3) Statistical Test, 4) A Linear Regression Analysis, and 5) Goodness of Fit and Cross-Tabulation. Chapters 1 through 4 provide 10 quizzes each, while Chapter 5 provides 30 quizzes. Students having at least 42 correct answers pass and receive a certificate. This course has been created with the cooperation of the Japan Statistical Society and Biometric Society of Japan (Takemura, et al., 2015; Izumi and Sue, 2017). The playback time of video lectures, scores of quizzes, and survey answers before and after courses are recorded in the Shiga University version of these JMOOC courses.

PBL practices

Students accumulate their experiences in real-data-oriented PBL practices from the first year of college onward and learn various aspects of problem-solving. They tackle exercises using the real data received from collaborative institutes and companies and public big data such as e-Stat (Figure 2) and Regional Economy Society Analyzing System (https://resas.go.jp/#/25/25202). By learning these practices for four years, students will acquire not only expertise in data analysis but also extensive knowledge in various fields.

In the course titled "Fundamentals of Statistical Analysis," we set up a theme of PBL practices and created educational materials in cooperation with the Disaster Prevention Research Institute, Kyoto University. As part of this course, a mini project is allotted to a group of five students to analyze data on climate-, earthquake-, and geography-related regional disaster prevention. The results of PBL practices are summarized in a report and a statistical graph poster. Such activities aim to enhance students' knowledge of statistics and emphasize the report and graph techniques of statistical expression. Following a poster tour, a group presentation is evaluated by teachers and the students of other groups, whereas a report is evaluated by a teacher alone. Rubric SPART is created based on three viewpoints: statistical literacy, reasoning, and thinking (Fukazawa, et al., 2018). Based on the rubric SPART for peer review, we create and use a checklist for each phase of the PPDAC cycle. The results of the evaluation are instantly fed back to students through an electric portfolio to encourage students to reflect on what they have learned.

Other courses on PBL practices use consumption and purchasing, annual health examination, social networking service, public statistic, and regional mobile data provided by collaborators. Topics chosen by students include a trend of purchasing chocolates, the relation between blood pressure and structure, Twitter data analysis for sport events, and a comparison of sex- and age distribution of visitors between cities.

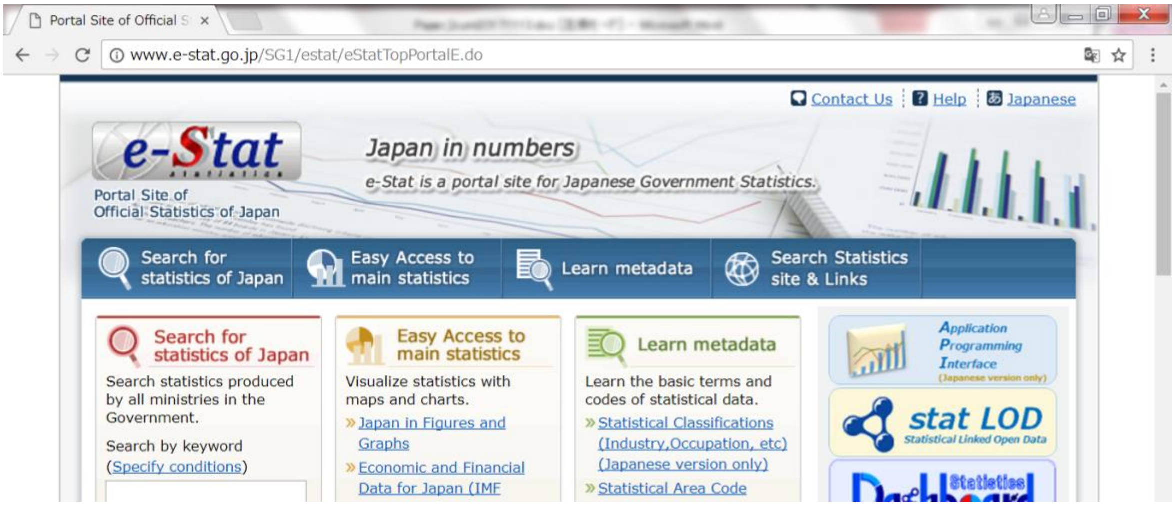

Figure 2. The first page of e-Stat (http://www.e-stat.go.jp/SG1/estat/eStatTopPortalE.do)

Collaboration with experts

Collaborating with experts plays a key role in the development and execution of our proposed model. The courses titled "Introduction of Data Science Case Studies and Hand-On Seminars" are taught by IBM Japan (IBM, 2017); Dai-ichi Life Holding, Inc.; the Data Scientists Society; and so on. Students obtain valuable experience in analyzing real data sets using commercial applications under expert supervision. In addition, the Division of Academic Affairs, Data Processing Center, and Support Office for Learning and Education on campus play an important role in the execution of our proposed model. They provide faculty training programs such as tutorials on LMS, interactive teaching, and learning disabilities. They prepare prints of handouts and statistical posters, as well. Moreover, they offer consultation services for students who experience difficulties in attending classes.

Comparison with other universities

Our educational model is consistent with the models prevalent in other countries. Our model is based on the Guidelines for Assessment and Instruction in Statistics Education college reports 2016 (ASA, 2016a). Our data science program includes a series of courses in data analysis and information science. When building our curriculum, we acknowledged the Curriculum

Guidelines for Undergraduate Programs in Statistical Science (ASA, 2014) and Curriculum Guidelines for Undergraduate Programs in Data Science (ASA, 2016b). We find some similarities between the undergraduate programs in statistics and data science program offered by Yale University (2017) and Ohio State University (2018). According to Coursera MOOC catalogue, the courses "Exploratory Data Analysis" and "Statistical Inference" offered by Johns Hopkins University (2018) seem to have similar contents as the courses "Statistics I: Basics of Data Analysis" and "Statistics II: Methods for Statistical Inference" used in our model.

Since our data science program has just completed its first year and we have no information regarding its performance during the coming years, it will be interesting to thoroughly examine the features in our model that do and do not work well after the first four years.

CONCLUSION

Our model is structured around an integrated combination of lectures in JMOOC, problem exercises, and small-group PBL practices. An interactive teaching method requires several hours of preparation in advance; the use of LMS and MOOC reduces this burden. In addition, it is important to maintain the motivation of students to learn outside the classroom and increase their interests in PBL practices. In this respect, choosing a familiar topic among students may be helpful. Furthermore, it is important to appropriately evaluate the results of implementation and instantly provide feedback to the students. By reviewing the merits and demerits of applying this model, we can improve our proposed model further. In the future, we will attempt to develop virtual teaching materials and textbooks, as well.

From a preliminary review of our model based on first-year experiences, we tentatively conclude that our model is a promising one from the perspective of future data science education. While students are satisfied with traditional lectures inside the classroom, they find it challenging to learn mathematical and statistical materials using MOOC. PBL practices motivate them to think about the possibilities of data science applications. Students may also recognize a reason why they want to become data scientists.

We hope that our proposed model will provide students with a more balanced introduction to statistical concepts, methods, and theory. Our model may encourage both science and liberal arts major students to consider statistics as a potential career path or, at the very least, a topic of high interest. Further, our model may guide teachers to employ student-centered pedagogy in their future classes.

ACKNOWLEDGEMENTS

We thank Professor Akimichi Takemura (Chief, Department of Data Science, Shiga University) and our colleagues for their supports. We also thank three anonymous reviewers for their many helpful comments. Partially supported by national university reform improvement promotion subsidies from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT), by Grant-in-Aid for Scientific Research from the MEXT (17K00047), and by the Institute of Statistical Mathematics (ISM) Cooperative Research Program (2017 ISM CRP-4306 and CRP-4308; 2018ISM CRP-4105 and CRP-4106) to the first author (S.I.).

REFERENCES

American Statistical Association. (2014). Curriculum Guidelines for Undergraduate Programs in Statistical Science. www.amstat.org/asa/education/Curriculum-Guidelines-for-Undergraduate-Programs-in-Statistical-Science.aspx

American Statistical Association. (2016a). Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Reports 2016. www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf

American Statistical Association. (2016b). Curriculum Guidelines for Undergraduate Programs in Data Science. www.amstat.org/asa/files/pdfs/EDU-DataScienceGuidelines.pdf

Fukazawa, H., Sakurai, N., & Izumi, S. (2013). Problem-based statistical education and its assessment in schools. Statistics Education and Research: Institute of Statistical and Mathematics, 275, 62-65. (in Japanese)

Fukazawa, H., Sakurai, N., & Izumi, S. (2018). Statistical Inquiry Process and Assessment. to appear in *Proceedings of the Institute of Statistical Mathematics*, *66*. (in Japanese)
IBM. (2017). IBM Watson. https://www.ibm.com/watson/jp-ja/
Iyoku, M. (2017). A strongest learning method to grow their own. SOGOU HOUREI Publishing, Co., LTD. (in Japanese)
Izumi, S., & Gould, R. (2010). International comparison of statistics education in undergraduate and graduate schools. *Statistics Education and Research: Institute of Statistical and Mathematics*, *243*, 43-46. (in Japanese)
Izumi, S., Sakurai, N., & Fukazawa, H. (2016). Interactive class design and its assessment in undergraduate statistical education. *Statistics Education and Research: Institute of Statistical and Mathematics*, *362*, 5-10. (in Japanese)
Izumi, S. (2017). Pilot study of problem-based learning practices (PBL) using public big data. *Statistics Education and Research: Institute of Statistical and Mathematics*, *379*, 97-102. (in Japanese)
Izumi, S., & Sue, M. (2017). Development of Introductory Statistical Courses in the Japanese Massive Open Online Course: A Preliminary Report. USCOTS 2017, Pennsylvania State University, May 18-21, 2017. Retrieved from www.causeweb.org/cause/uscots/uscots17/program/posters
Johns Hopkins University. (2018). Coursera MOOCs. www.coursera.org/specializations/jhu-data-science?utm_source=gg&utm_medium=sem&campaignid=313639147&adgroupid=35684912840&device=c&keyword=executive%20data%20science%20coursera&matchtype=b&network=g&devicemodel=&adpostion=1t1&creativeid=242839477915&hide_mobile_promo&gclid=EAIaIQobChMIrfXjkYru2AIVhbbACh1CfwOoEAAYASAAEgIZPPD_BwE
Maruyama, H., Kamiya, N., Higuchi, T., & Takemura, A. (2015). Developing data analytics skills in Japan: status and challenge. *Journal of Japan Industrial Management Association*, *65*, 334-339.
Ohio State University. (2018). Undergraduate data analytics major. data-analytics.osu.edu/
Sakurai, N., Izumi, S., & Fukazawa, H. (2015). Aspects of certification in Massive Open Online Courses (MOOC). *Statistics Education and Research: Institute of Statistical and Mathematics, 335*, 73-78. (in Japanese)
Shiga University. (2017a). Data science view, Shiga University.Vol.1. Hikone: The Center for Data Science Education and Research, Shiga University. (in Japanese)
Shiga University. (2017b). Curriculum policy in Faculty of Data Science, Shiga University. http://www.shiga-u.ac.jp/information/info_public-info/public-education/public-education_hikone/ (in Japanese)
Shiga University. (2017c). Website of Faculty of Data Science. https://www.ds.shiga-u.ac.jp/
Takemura, A., Sakaori, F., Nakayama A., & Shimokawa, T. (2016). Statistics I: Basis of Data Analysis, Official Study Note (2nd Edition). Tokyo: Japanese Statistical Association. (in Japanese)
Takemura, A., Shina, Y., Izumi, S., Matsuda, Y., & Sato, T. (2015). Statistics II: Methods for Statistical Inference, Official Study Note. Tokyo: Japanese Statistical Association. (in Japanese)
Takemura, A., Izumi, S., Saito, K., Himeno, T., Matsui, H., & Date, H. (2018). Shiga-University Model of Data Science Education. to appear in *Proceedings of the Institute of Statistical Mathematics*, *66*. (in Japanese)
Yale University (2017). Undergraduate Statistics and Data Science Programs of Study 2017–2018. catalog.yale.edu/ycps/subjects-of-instruction/statistics/