Use of the R Commander in the Introductory Social Statistics Course: A Case Study

Takashi Yamashita and Rachel Caitlyn Crane

Abstract—Statistics anxiety is prevalent among social science students in higher education. One of the known strategies to reduce statistics anxiety is to provide more opportunities for the hands-on training to practice statistical skills. This case study incorporated the open source statistical software with the graphic user interface—R Commander—to the existing social statistics course and assessed the changes in the students’ attitudes toward statistics using the pre- and post-test design in one semester. Results (n = 29) showed that one of the attitudinal domains (i.e., affect) significantly improved whereas other domains did not change. Also, the observed issues recorded in the field notes were summarized and practical solutions were discussed. Use of an open source software is a cost-effective strategy to increase the hands-on training in social statistics course in hopes to alleviate the statistics anxiety and to promote lifelong statistics learning among social science students.

Index Terms—Social statistics, statistics anxiety, sociology, R commander.

I. INTRODUCTION

Quantitative literacy is one of the six essential competencies set by the American Sociological Association (ASA) [1]. Thus, sociology majors should be trained in becoming sophisticated consumers of social statistics as a future full member of the society [2].

Yet, according to the ASA, the quantitative literacy gap has been a persistent issue in undergraduate sociology programs over the last several decades [3]. Sociology students generally find the statistics more difficult than other subjects [4]. In fact, carrying out statistical analysis using software and interpreting the results are some of the most common skills in which sociology students lack confidence [5], [6]. To make matters worse, statistics is also known to be a source of anxiety (i.e., statistics anxiety) [6]-[8]. This statistics anxiety is often triggered by general math anxiety or “math-phobia” due to the greater emphasis on mathematical aspects of statistical methods than the statistical concepts [9].

Sociology programs should care about educating and training their students in both statistics and quantitative literacy for several reasons. First, quantitative literacy is necessary to fully understand the empirical basis of sociology as an academic discipline [3]. Second, quantitative literacy is an essential skill to becoming informed citizens equipped with appropriate statistical thinking and reasoning skills needed in our data-rich society [4], [7]. Third, growing volume and complexity in social science data suggest a greater importance for academics to develop the ability to understand and speak complex statistical information. In this respect, social science students are well-positioned to connect numerical data to real-world social issues [10]. Fourth, students tend to see the required statistics course and their statistics anxiety as barriers to completing the degree program and may choose other majors [8]. Last but not least, statistical knowledge and quantitative analysis skills are linked to the employability because they are highly valued by potential employers [11], [12]. Howery and Rodriguez (2006) argue that most sociology programs are aware of the value of statistics and research skills for student’s professional development towards future success.

Despite the fact that the students’ statistics knowledge gain is one of the primary goals of statistics training in the sociology program [13], one may argue that taking a statistics course or two does not achieve sufficient quantitative literacy. In other words, learning statistics should be a part of lifelong learning process, and statistics training in the sociology programs should build the foundation for further learning. As such, reducing statistics anxiety in addition to enhancing student motivations/interests in quantitative literacy are also important aspects of the statistics education for sociology undergraduate students [14]. At the least, the required statistics course should not deteriorate students’ attitudes toward statistics.

In the literature, two strategies for alleviating statistics anxiety (and in turn to promote positive attitudes toward statistics) have been suggested. Several studies reported that greater exposure to statistical problems, particularly through hands-on training, effectively reduce statistics anxiety in social science programs [15]-[19]. On a related note, Forte (1995) stresses that the use of real-world data in the context of specific sociological problems is key for successful applied statistics training. Also, statistics education experts suggest focusing on the conceptual understanding and statistical reasoning rather than the mathematical aspects of statistical techniques (e.g., underlying mechanics, formula, equations) in social science programs [9], [20].

In this respect, use of statistical software is critical in social statistics courses to reduce the amount of computational/mathematical work by hand, and in turn, to alleviate the statistics anxiety [21], [22]. However, there are several barriers to using the statistical software as a means of increasing the amount of hands-on training and reducing the computational burden. A comparatively limited number of statistical software with intuitive graphic user interfaces (GUI: point-and-click operation) are currently available. For
instance, IBM SPSS [23] and STATA [24] feature user-friendly GUI. Meanwhile, other predominantly syntax-based software [e.g., SAS [25] and R [26]] are most likely too challenging for use in undergraduate social science students due to the necessary syntax and programming skills [7, 10]. Also, it is widely known that the cost of commercial statistical software (e.g., institutional and individual license) is of a concern not only to the educational institutions but also to the students. Some students may have poor access to the software due to the available lab computers nearby and/or incompatibility with their personal computers (e.g., Chromebook and Linux-based computers). Other barriers include that students may face difficulty getting help from instructors in a timely manner due to the poor access to the software and/or instructor’s office hours. The barriers could be even more serious when the class size is large.

Taken together, there is a need to increase the hands-on training to address the quantitative literacy gap and to promote positive attitudes towards statistics among social science students. As one of many possible strategies, we employed an additional level of hands-on training using the open source application — R [26] and its GUI application — R Commander [27] and added it to our existing introductory statistics course with SPSS. We assessed the students’ attitudes toward statistics using the pre- and post-test design. R Commander (described in more detail later) is a suitable option since it is both free of cost and equipped with the GUI to avoid the steep learning curve often seen in the syntax-based applications. We did not replace our conventional software (i.e., SPSS) to maintain consistency with our introductory methods course and previous departmental software preference. Our research question asked if the additional R Commander hands-on training in the existing introductory statistics course with SPSS changes students’ attitudes toward statistics in the undergraduate sociology program.

II. DATA AND METHODS

A. Course Description

Study was conducted in an introductory statistics course in the sociology undergraduate program in a mid-size, U.S. state university with roughly 14,000 undergraduate and graduate students on the East coast. Our university is classified as one with higher research activity by the Carnegie Foundation [28]. The introductory social statistics course is one of the required courses in the undergraduate sociology program. This course covers the basic statistical concepts (e.g. probability, levels of measurement, sampling) in addition to descriptive and inferential statistical techniques to analyze the social scientific data. To enroll in this course, students must complete the introductory sociology as well as sociological research methods courses with the minimum passing grade of “C” and the statistics readiness test for the basic mathematical and algebraic operations. This four-credit-hour course meets twice a week — lecture and lab sessions — in a regular semester (an hour and forty-five minutes each session). Conventionally, SPSS alone has been used for the lab session. The lab session is based on the contents covered in the lectures and SPSS exercises in the textbook [29] although the textbook may vary by the instructors. The lab exercises mainly employed the publicly available social data from the latest wave of the General Social Survey (GSS) [30]. One section of this course is usually offered in the spring and fall semesters. Two to three instructors take turns teaching this course.

In the present study, 47 students registered for the introductory social statistics course. The number of students was greater than usual and the available computer lab was not able to accommodate this course. As such, we requested the students to either install the SPSS and R/R Commander on their personal computers or use the university’s virtual desktop/machine through the web browser or desktop client software. Access to the SPSS license is not currently available for free to the students at our university. Based on an informal in-class assessment, about half of the students had both SPSS (purchased by students) and R Commander installed on their own computers. Others used the virtual desktop for the SPSS, R Commander or Both. The tenured faculty and a teaching assistant with graduate-level statistics training were the instructors in the current study.

B. R Commander

R Commander [27] is one of the packages in R [26]. R Commander provides the GUI for the basic statistical analysis and allows point-and-click approach in the R environment. R Commander is available at no cost and compatible with the Windows, Mac and Linux operating systems. R Commander is not included in the base R and therefore, the package needs to be downloaded from the Comprehensive R Archive Network (CRAN) and installed. Students need to use two simple syntaxes—install.packages("Rcmdr") — and — library(Rcmdr) --- to download/install the packages, and launch the application in R, respectively. The R syntax is case-sensitive. R Commander has the built-in functions to import data in a variety of formats such as SPSS, SAS, STATA and Microsoft Excel. The data manipulation (e.g., recoding variables, reassigning the level of measurement) is possible directly in R Commander. Also, R Commander can generate descriptive statistics and conduct a series of statistical significance tests. Moreover, data graphics can be produced using the point-and-click features. One of the advantages of R and R Commander is the rich online repository of information and numerous help forums/communities where students may find solutions to virtually any common issue. More detailed descriptions, practical guidance, and applications are published elsewhere [31], [32].

C. Pre-test and Post-test Instrument

Attitudes Toward Statistics (SATS-36) [33] was used to assess the students’ attitudes toward statistics in the six attitudinal domains including affect, cognitive competence, value, difficulty, interest, and effort (see more detail in Table 1). The first four domains have been validated and had acceptable internal consistency (Cronbach’s alpha >= 0.64). However, the interest and effort domains have not been tested for the psychometric properties. The SATS-36 consists of the pre-test and post-test. In the present study, the data
were collected in the first two weeks (pre-test) and last two weeks (post-test) of the spring semester in 2018. Students received the credit (about 2% of the total grade) for the pre- and post-test when the student self-reported their completion. Students were invited to complete the pre- and post-test in an online survey platform [34]. The survey included the informed consent, demographic information, prior statistics software experience and software preference in addition to the SATS-36. The unique ID based on the day of birthday, last 2-digit of the current ZIP code, and last 2-digit of the primary phone number were created to match the pre- and post-test survey. The use of SATS-36 for this case study was permitted by the developer [33]. The survey instrument and study protocol were approved by the Institutional Review Board of the “Authors’ institution” (protocol #[masked for the blind review]). Additionally, we kept the field note to record any issues that we recognized throughout the semester.

D. Analytic Plan

All analysis was conducted using the R version 3.4 [26]. The descriptive statistics are computed for the demographic characteristics, statistics software experience/preference and indices for each of six attitudinal domains including affect, cognitive competence, value, difficulty, interest, and effort in SATS-36 [see 33 for the individual survey items]. Each domain consists of 4 to 9 survey items and the average is used as a summary index. Acceptable validity and reliability have been reported for the first four domains using the diverse samples of the college students [33]. In this study, the reliability of each domain is assessed by the McDonald’s omega and Cronbach’s alpha coefficients which were estimated with the psych and GPArotation packages in R [35]-[37]. The omega and alpha coefficient greater than 0.60 was considered acceptable. The preliminary power analysis using the R pwr package [38] showed that the minimum sample size for the alpha = 0.05, 1-beta = 0.8 and medium to large effect size (d = 0.5 – 0.8) was 33 – 14. Also, all attitudinal domains showed a skewed distribution. Therefore, the pre- and post-test comparison was conducted using the Wilcoxon signed-rank test instead of the conventional dependent t-test [39].

III. RESULTS

During the pre- and post-tests data collection periods, 38 students attempted at least one of the surveys after excluding 4 registered students who did not participate in this course and 5 students who did not complete the majority of survey or did not report the key information (e.g., unique ID). Of those, 29 (76% of 38) students completed both the pre- and post-tests. The average age of the respondents was 24.90 years old (Standard deviation = 7.11). About 55% was female and 72% were in the senior year of the program. The post-test results showed that about 66% of students reported the neutral or positive attitudes toward the SPSS whereas 59% of them reported the neutral or positive attitudes toward the R Commander. Additionally, 65% and 41% of students reported that SPSS and R Commander helped their statistical learning, respectively. In terms of software preference, 41% of the students reported that they would like to use the R Commander only or both R Commander and SPSS if they were to re-take this course. Yet, 58% preferred only SPSS.

Results from the reliability analysis and Wilcoxon signed-rank tests were shown in Table I. For both the pre- and post-tests, the affect and interest domain showed the consistently acceptable reliability. At the same time,
cognitive competence, value, difficulty, and effort had partially questionable reliability accordingly to the omega coefficients being less than 0.60 although the alpha coefficients were greater than 0.60. Given the small sample size, we did not attempt to re-scale the SATS-36 and decided to interpret the results with caution for the domains with questionable reliability. There was a statistically significant difference in the affect domain between the pre- and post-tests ($V = 87, p < 0.05$). Specifically, the “student’s feeling concerning statistics” became significantly more positive during the semester. However, there was no statistically significant change in other five domains.

Fig. 1. Sample instruction of the R and R commander set-up.

IV. DISCUSSION

Teaching statistics is critical to train social science students to become citizens with sufficient quantitative literacy in information-rich societies. Yet, social science students frequently see statistics as an obstacle and source of anxiety. To test a relatively unique teaching design — use of an open-source statistical software — to increase statistics hands-on training to make positive changes in the attitudes toward statistics in the sociology program, this study incorporated the R Commander exercises into the existing statistics course at the mid-size public university. The student attitudes — affect — have significantly improved across the length of the semester although there was no significant change in other attitudinal domains. It should be noted that the small sample size might not have captured the minor change (i.e., small effect). Also, the relatively short time span of one semester or the limited amount of additional hands-on training in this course might not have been sufficient to make sizable changes in some of the domains.

Over 40% of students reported the learning benefit from the R Commander exercises although more than half of the students preferred SPSS. Our analysis showed that the additional hands-on training with R Commander was linked to more positive attitudes toward statistics (i.e., affect domain). In addition, results also suggested that including additional hands-on training with R Commander did not negatively impact students’ attitudes toward statistics. As such, we argue that improved attitudes toward statistics, known benefits of R Commander, and no negative impact on the students’ attitudes toward statistics collectively justify the use of R Commander in future social statistics courses.

Results from this study were generally consistent with the previous studies reporting students’ preference of SPSS, which provides a little if any, benefits to their attitudes toward statistics [7], [40]. In this study, even with the additional R Commander component, only one (i.e., affect) of six attitudinal domains showed a significant change. Nevertheless, the empirical research on the attitudes toward statistics among social science students is still limited. Since our study was a case study with relatively small sample size, additional research with larger and more diverse samples should verify the results.

A. From the Field Note

Here, we share some of the key issues from our field note and possible solutions. First, students who needed or wanted to use the virtual desktop application spent somewhere between 5 and 15 minutes to set up the SPSS and R Commander. Since R in the virtual desktop did not have the R Commander package, students needed to download and install the package for every lab session. Arranging with the
IT department to ensure they installed R Commander for all virtual desktop machines would have saved the class time. That time could have been spent completing the exercises and utilizing statistics techniques. Relatedly, several students struggled to set up the R Commander throughout the semester using the syntax even though a step-by-step instruction was provided in multiple formats (e.g., handouts with screenshots, online resource, video instruction: see Fig. 1 for an example). The instructors and teaching assistant may need to repeatedly demonstrate how to set up R Commander. Also, we have encountered the compatibility issues of R and R Commander. For example, when the newest version of R is installed, R Commander occasionally did not launch. The instructor had a few students re-install an older version of R to successfully launch R Commander.

Additionally, the perception toward the additional hands-on training was roughly split between two opposing sides. Some students saw it as a learning opportunity while others saw it as an additional burden. It is important for instructors to make the objectives of additional components clear at the beginning of the semester. Moreover, when multiple instructors teach the same course in a different section or semester, intradepartmental coordination and prior agreement of the software choice is indispensable to maintain consistency as a program. Furthermore, we believe that instructors and graduate students with a basic statistics background could quickly learn how to use R Commander although the instructors may need to learn themselves (if they lack prior experience) and to train the teaching assistant in advance. Furthermore, after several attempts at the beginning of the semester, we realized that having students replicate the SPSS exercise with the R Commander instead of giving them new problems was a better strategy to engage students. As a result, students seemed to be able to practice reading the same results/outputs in two different formats. If we were to exclusively use R Commander for a new problem and/or data management/manipulation, it would most likely require additional course preparation, demonstration, and individual-assistance merely for the software use rather than focusing on the statistical concepts and interpretation of the results [21]. Finally, although more than half of the students did not like the R Commander exercises in this study, student evaluation or course satisfaction may not be the best indicator of statistics education outcomes [13]. Different evaluation strategies such as a completion of complex data analysis tasks instead of the traditional exams may not only show the progress of students but also reduce statistics anxiety [6], [16].

B. Limitations

Several limitations of this study should be noted. To being with, this study was a case study with one typical introductory statistics course in the mid-size state university in the U.S. Since generalizability was not a part of the study design, additional research is required to verify the findings. Also, according to the McDonald’s omega coefficients, four out of six attitudinal domains in the SATS-36 showed the questionable reliability in this study despite the previously reported sound psychometric properties [33]. At the same time, the Cronbach’s alpha coefficients consistently showed the acceptable reliability in this study. The small sample size, as well as the unknown underlying factor structure, might have contributed to relatively low reliability in four domains [35]. Future research should examine the underlying factors and psychometric properties of the SATS-36 with larger and more diverse samples of sociology/social science students. Additionally, the findings from this study should not be translated into any causal inferences as additional studies are needed to verify the direct link between additional hands-on training and attitudes toward statistics in more rigorous experimental design (particularly one using control groups). Finally, this study did not evaluate the quality of the R Commander exercises provided in the lab session. In other words, the instructor was solely responsible for designing the R Commander exercise in this course. It may be useful to have a different instructor evaluate those exercises to maximize the benefits. Perhaps, qualitative inquiry with students is also useful to identify space for improvement. Nonetheless, increasing the hands-on training is generally linked to lower statistics anxiety [e.g., 16].

C. Practice Implications

Despite the limitations, based on the results from this study and insights from the previous research, two preliminary practice implications can be drawn. First, providing additional applied exercises to an existing statistics course using the open source software increases practice opportunities, although it requires a small amount of additional investment (e.g., instructor’s time and training). One of the known strategies to reduce the prevalent statistics anxiety and in turn, to promote the attitudes toward statistics among the social science students is to give students more opportunity for hands-on training [15], [17]. This study found that the statistics course with additional R Commander exercises yielded a positive impact on one of the chosen attitudinal domains while not negatively impacting the students’ attitudes toward statistics. Second, an assessment of students’ attitudes toward statistics in social science programs should be regularly conducted in hopes to promote the lifelong quantitative literacy learning. As an academic program, statistics instructors as well as department program coordinators may discuss and unify the overarching goals (e.g., teaching applied statistical skills and reducing the statistics anxiety). When a new teaching strategy such as adding the R Commander exercise is implemented, monitoring students’ attitudes towards statistics is beneficial while evaluating their skill/knowledge development [13]. Assessments in statistics education in general and that in the social science programs, in particular, is still scarce [9].

V. CONCLUSION

This case study provided preliminary empirical evidence for the use of open source statistics software — R Commander — as an additional hands-on training in the introductory statistics course in the sociology program at the mid-sized state university in the U.S. Although more than half of the responding students did not like the R Commander exercises, over 40% of students reported the learning benefits. Importantly, the inclusion of R Commander exercise in the
course not only partially improved students’ attitude toward statistics but also made no negative impact on the students’ attitudes during the semester. Considering the relatively small amount of preparation required, we argue that additional hands-on training using R Commander is feasible and beneficial to statistics courses in the social science programs. Building on the findings from this case study, future research should incorporate R Commander hands-on training into courses and to conduct a program-wide long-term assessment of students’ attitudes toward statistics to facilitate innovative statistics education for social science programs at the higher education institutions.

ACKNOWLEDGMENT

The authors report no conflict of interest in this research. The authors report that this research has received no funding from any agency or organization.

REFERENCES


Yakisshi Yamashita is from Tokyo, Japan. He earned his bachelor’s degree from Tokyo Gakugei University in 2003. He earned two master’s degrees in the applied gerontology and sport administration at Ball State University, Muncie, Indiana, USA in 2007. He completed his doctoral training in the Department of Sociology and Gerontology at Miami University, Oxford, Ohio, USA in 2011. He has one-year post doctoral training at the Scripps Gerontology Center of Miami University. In 2017, he completed his master’s of public health program at University of Nevada, Las Vegas. He is currently an associate professor of sociology and core faculty of gerontology Ph.D. program at University of Maryland, Baltimore County, Baltimore, Maryland, USA. Previously, he was an assistant professor of sociology at University of Nevada, Las Vegas between 2012 and 2017. Dr. Yamashita is a member of the Gerontological Society of America. He is an active editorial board member of the Journal of Gerontology Social Science.

Rachel C. Crane was born in Baltimore, Maryland on September 28th, 1994. She earned her B.A. in both social work and sociology from University of Maryland, Baltimore County in 2016, and her master’s degree in applied sociology from that same institution in May 2018. From 2013-2016 she worked as a peer mentor assisting the application of her university’s redesigned
SOCY 101 course. Between 2016 and 2018 she worked as a teaching assistant for the sociology department’s research methods and statistics courses. She is currently working at the University of Maryland, Baltimore county as a part-time lecturer in the Sociology Department, covering the Introduction to Sociology course, as well as a psychiatric rehabilitation program assistant and electronic health record trainer at an outpatient mental health clinic in Windsor Mill, Maryland.

Prof. Crane is a member of the Iota chapter of the Alpha Kappa Delta honor society. In 2018 she was the recipient of the Gill Award for students in her department and discipline. She also earned a Certificate in the Non-Profit Sector during her graduate work at her university. In February of 2018 Prof. Crane presented her master’s paper “The Impact of Family of Origin on the College Experiences of Traditional-Aged, First-generation College Women” as a poster at the Eastern Sociological Society conference in Baltimore, Maryland.