# USE OF COMMERCIAL AND FREE SOFTWARE FOR TEACHING STATISTICS

BRYSON STEMOCK
*Youngstown State University*
*bjstemock@student.ysu.edu*

LUCY KERNS
*Youngstown State University*
*xlu@ysu.edu*

## ABSTRACT

*This study investigates the effectiveness of two statistical software packages, the commercial software, SPSS, and the free software, R Version 3.5.0, in teaching statistics, by comparing grades earned by students in an introductory statistics course. Students taught using R earned slightly higher grades overall than those in the SPSS class, although the difference between the two groups of students was found to be not significantly different. The responses given by students on the end of semester survey for the use of technology indicated that students felt comfortable with both software packages, and the software R was recommended more strongly for future use in a statistics course than the software SPSS. These findings provided some insights on the use of different software packages in statistics education, which might support the use of free software in teaching statistics at college level.*

***Keywords:*** *Statistics education research; Statistical software; R; SPSS*

## 1. INTRODUCTION

Statistics, as the science of collecting, analyzing, and interpreting empirical data, provides tools for making informed decisions based on data and plays a vital role in every field of human activity and every discipline. For many undergraduate students in college, statistics is a difficult subject and is often perceived as the major hurdle standing in the way of earning a college degree. Many studies have shown that the use of statistical software can make teaching and learning statistics more effective as it encourages active learning, enhances students' understanding of statistical concepts, and improves their problem-solving skills (Garfield, 1995; Gomez, 2010, 2011; Higazi, 2002). The recent Guidelines for Assessment and Instruction in Statistics Education (GAISE) college report, revised in 2016 by the American Statistical Association, underscores the importance of technology in statistics education and recommends the use of statistical software to perform most computations to allow greater emphasis on understanding concepts and interpreting results.

There is a multitude of professional software packages available on the market (e.g., SPSS, SAS, R, DataDesk, JMP, Minitab) designed for the explicit purposes of performing statistical analyses. Additionally, many different types of educational software programs and applets have been developed exclusively for educational purposes, that is, for the sole purpose of helping students learn statistics (e.g., Fathom, TinkerPlots, OpenStat, Instat, InspireData). Among the existing statistical software packages, some are commercial products (e.g., SPSS, SAS, Minitab, InspireData), whereas others are free software programs (e.g., R, OpenStat, SAS University Edition). The choice of software programs to teach in a statistics class can have a great impact on the quality of education—the use of suitable software in teaching statistics can facilitate and improve the learning of statistical concepts as well as help students build software skills needed for their future careers. Therefore, although the benefits of using technology in statistics education are evident, it also raises a number of important issues to

*Statistics Education Research Journal, 18(2), 54–67, http://iase-web.org/Publications.php?p=SERJ*

consider when selecting appropriate technology to use in a statistics class (Biehler, Ben-Zvi, Bakker, & Makar, 2013; Chance, Ben-Zvi, Garfield, & Medina, 2007).

One of the main issues is the effectiveness of the technological tool in aiding instruction and learning in statistics; that is, whether the software fits students' particular area of study and the use of technology can successfully enhance their learning of statistics. There are a number of factors to consider when deciding the appropriate software that matches the learning and teaching needs. These include ease of use (i.e., ease of data entry, ease of data processing, etc.), output presentation, interactive capabilities, portability, and flexibility of the technological tool, among others (Chance et al., 2007; Franklin & Garfield, 2006). Regarding ease of use, it is crucial to consider the academic background of the students; for example, students who major in the science, technology, engineering, and mathematics (STEM) fields usually have stronger programming skills than non-STEM students, and hence are bettter prepared for the use of a command line interface software program such as the R program. If students are not ready or don't have the necessary background for some sophisticated software packages, then use of these programs may result in students spending more time learning to use the software than applying it. Factors such as portability and flexibility can influence students' perceived usefulness of the software product, which is a strong determinant of students' intentions to use the tool. Students tend to perceive a software package as useful if it can be utilized in other courses, such as using the same technological tool to carry out other tasks of data analysis or simulation, and hence will be motivated to learn the technology. In this regard it is also important to consider students' future studies when choosing suitable software.

Another important issue is the availability and affordability of the technology. Traditionally, commercial statistical software like SPSS or SAS has been used as the standard software in college level statistics courses. Commercial software such as SPSS is expensive (Paura & Arhipova, 2012), and at most institutions is not provided for students or individual users. Many colleges and universities own a site-license for the commercial software, and the licensed software is often installed on lab computers, which makes it convenient for students to work on their coursework on campus. A problem may be created, however, for students who complete course related work off campus, particularly for those who live off campus. Because of the high costs, students often choose to purchase a cheaper student version for home use: the student versions do not match the full versions provided by the university, and are sometimes inadequate for carrying out some data analysis tasks, usually more complex and advanced tasks, because the student versions are often stripped of certain essential features for peforming those analyses. In this respect, using free and open source software such as the statistical software package R seems to be a preferable alternative to commercial software.

Popularity of software within the job market is another issue of growing importance to consider when selecting suitable software. As technology has fundamentally changed every aspect of the way many businesses operate in our society, the need to choose statistical packages that are aligned to the skills required by employers is becoming more evident—an appropriate software choice can provide students with essential software skills that many employers are currently seeking and will continue to look for in the future. Ozgur, Kleckner, & Li (2015) discussed the relative popularity of various software programs based on the survey results provided by Muenchen (2014), and they noted a discrepancy between data analysis software used and software sought within the job market. Ozgur et al. conducted a follow-up survey on the software taught by faculty in 18 departments to understand the potential disconnect between the software skills that are taught at school and sought by employers. The results (Table 1, Ozgur et al., 2015, p. 3) indicated that a wide variety of software was taught by universities and particular software choices varied among departments and universities; for example, for small, Midestern, private universities and undergraduate courses, software choices were SAS and Excel in the Department of Mathematics & Computer Science, Minitab and GRETL (Gnu Regression, Econometrics and Time-series Library) in the Department of Economics, and R in the Department of Statistics.

As more and more statistical software packages become available, it is becoming increasingly important to conduct coordinated research that specifically addresses these important issues related to the use of technology, particularly on the comparative effectiveness of different software when teaching statistics. In spite of this urgent need, it appears that very few research studies have been reported in the literature specifically assessing the comparative advantages and disadvantages of competing statistical software. Chen, Moran, Sun, & Vu. (2018) conducted a study to assess the effectiveness and efficiency of two statistical software programs, Intellectus Statistics (IS, a cloud-based statistical analysis software) and SPSS. The participants (students at California State University) were asked to complete some statistical data analysis tasks as well as a post-test questionnaire for perceived usefulness and ease of use of the two software programs. The authors did not find significant differences between the two software on accuracy scores or time on tasks; however, the responses from the post-test survey indicated that the participants preferred IS over SPSS and felt that IS increased their confidence in performing statistical analyses. Wang, Chen, Schifano, Wu, & Yan (2016) focused on the challenges involved in statistical analysis with big data and discussed recent software developments (R, SAS, SPSS, and MATLAB) that specifically address the big data challenges. Paura and Arhipova (2012) explored the advantages and disadvantages of commercial (SPSS, SAS, and Excel) and free software (R) for teaching statistics. Other published literature includes Proctor (2002), Wilhelm (2007), Nunes, Alvarenga, Sant'ana, Santos, & Granato (2015), McNamara (2015), and Ozgur et al. (2015).

The present work intends to provide information to further our understanding of the role of different technology for effective teaching and learning of statistics by investigating the effectiveness of two popular statistical software packages, R and SPSS, on students' learning outcomes in an introductory statistics course. The primary concern of the study was to compare the effects of the two software programs by examining differences in academic performance measured by test scores earned by students, and to seek evidence that might support the use of free software in statistics courses at college level.

The study was conducted between August 2017 and May 2018 using an introductory statistics course in the Department of Mathematics and Statistics at Youngstown State University. The course selected for the study, Probability and Statistics, is an introductory statistics course with the primary goals of understanding the fundamental concepts and principles of statistical reasoning and achieving proficiency in statistical data analysis. The prerequisite for the course was the completion of both Calculus 1 and 2.

## 2. STATISTICAL SOFTWARE

### 2.1. SPSS

SPSS was first launched in 1968 as the Statistical Package for the Social Sciences. It was acquired and extended by IBM in 2009, and is now used well beyond the social sciences. Although it has retained the name SPSS, it is no longer an acronym, and the software is officially known as IBM SPSS (IBM SPSS statistics for windows, 2017). IBM SPSS is commercial software and a common tool used in teaching statistics at the college level. It is also one of the most widely used computer programs for the analysis of quantitative data in the social and behavioral sciences (Koh & Witarsa, 2003; Ozgur, Dou, Li, & Rogers, 2017). SPSS comes with a command-line interface and graphical user interface (GUI) with pull-down menus. There are three windows (Data Editor, Viewer, and Syntax Editor) in SPSS, and each window has a different role. The Data Editor window offers a simple and efficient spreadsheet-like utility for entering, editing, and browsing the data file. It has two distinct views, Data View and Variable View, which are accessible through the tabs at the bottom of the Data Editor window. The Data View displays the working data file and the Variable View shows the properties for each variable in the dataset. The Viewer window contains the results from statistical procedures and the Syntax Editor window allows the user to write SPSS commands.

The graphical user interface provided by SPSS offers a convenient way for users to see and select options on the drop-down menus (such as Files, Edit, View, Data, Analyze, etc.) for carrying out statistical analyses, and hence makes it very easy and intuitive for all levels of users. Another advantage of SPSS is that it allows users to import data from other sources (when data are organized as rectangular datasets, such as Excel spreadsheet) in a few basic and straightforward steps. SPSS suffers from certain limitations, however, including 1) SPSS is inflexible because it offers limited amount of options for data analysis and has poor programming tools; 2) SPSS often overburdens the user with too much output—the voluminous output confuses the user about on which parts of the output they have to focus, which often stipulates more questions that reach far beyond the user's knowledge of and experience with statistics; 3) SPSS is expensive.

### 2.2. R

R is a free, open source software package for statistical computing and graphics, based on the S language developed by Bell Laboratories. It was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand in 1991 and initially released in 1993 (Ihaka & Gentleman, 1996), and is currently developed by the *R Development Core Team* (R Core Team, 2017). R uses a typed language via a command line interface (CLI), and many statisticians and data scientists use R by entering commands at the prompt (> by default) in the CLI. For users who are not familiar with the standard command-driven interface and hence will consider it as an obstacle, a number of developments have been made to create a user interface which help flatten the learning curve or even free the users from learning command language; among them, RStudio and R Commander are two popular ones.

RStudio (RStudio Team, 2015) is a free, open source integrated development environment (IDE) for R that provides an alternative user interface to R. It organizes its interface to make it convenient for the user to view R code, source/data, plots, history and workspace simultaneously, and provides a number of features to make R easier to use, including an interactive console for using the R language and environment, an ability to import data files directly (CSV, Excel, SPSS, SAS, and other datasets), and support for writing R scripts with syntax highlighting, code completion, smart indentation, and interactive debugging. A free version of RStudio is available for download from https://rstudio.com/products/rstudio/download/. R Commander (Fox, 2005, 2017) is a graphical user interface for R, one that provides a point-and-click interface to R allowing users to focus on statistical concepts and methods rather than on learning and remembering R commands. Hence, it is a useful tool for users of R who find the standard command-line interface difficult to use. It is implemented as an R package, the Rcmdr package (Fox & Bouchet-Valat, 2018), which is freely available on the Comprehensive R Archive Network (CRAN, via https: //CRAN.R-project.org).

In R, a statistical analysis is normally done as a series of steps, and intermediate results are stored in objects. Therefore, whereas SPSS often gives copious output from a statistical analysis, R gives minimal output and stores the results in objects, which can be easily accessed and substituted into subsequent R commands. Because R is open source, it is easy for the user to inspect, modify, and enhance source code. In addition to the built-in statistics functions in the base R environment, R also provides many powerful analytical and graphical tools via *packages* (collections of functions and data sets developed by the R's user community), which makes it very easy to implement extremely useful "modern" statistical methods such as bootstrapping, permutation tests, simulation studies, and others. Base R includes "base" packages and the CRAN recommended packages, both of which are recognized as reliable (R Foundation, 2018). These packages are made available automatically when you install R; STATS (base package) and MASS (recommended package) are particularly useful for introductory statistics courses. Although there are numerous user-contributed R packages, which further extend the capabilities of R, users need to be aware that although many packages are of excellent quality, others may be less than perfect and the maintenance of packages depends on the efforts of the developer and R community members. Because R is a language, it offers power and flexibility, especially when it comes to handling highly repetitive tasks and complex statistics analyses that are not available through the menus. This can be particularly powerful and effective for people who have a good working knowledge of R programming (Valero-Mora & Ledesma, 2012).

## 3. METHODS

### 3.1. PARTICIPANTS

A convenience sample of 67 undergraduate students enrolled in the statistics course Probability and Statistics at Youngstown State University (a mid-sized comprehensive urban research university in Youngstown, Ohio) were selected for this study. Thirty-four students enrolled in the course in Fall 2017, and 33 registered for the course in Spring 2018. All students were taught using R in Fall 2017 and SPSS in Spring 2018. Most were upper-division students who were taking the statistics class to meet a major requirement for engineering, biology, mathematics, or computer science. Four students withdrew from the course in Fall 2017 and two students withdrew from the course in Spring 2018, and they were excluded from the study. The final data set contains 61 participants.

The Institutional Review Board of the University approved this study. Informed consent forms were given to all students to obtain permission for using de-identified performance assessments for this study and subsequent publications. Students were given the opportunity to ask questions about the study and were informed that participating in the study was not mandatory and would not affect their course grade.

### 3.2. DESIGN AND MATERIALS

This study was a 2 × 5 mixed-model design in which participants were divided into two groups (R group and SPSS group) and all participants were given five tests (pre-assessment, test 1, test 2, test 3, and final exam). Thus, in this study, software packages was a between-subjects variable with two levels (R vs. SPSS) and testing period was a within-subjects variable with five levels.

The required textbook for both classes were *Probability & Statistics for Engineers & Scientists* ($9^{th}$ edition) by Walpole, Myers, Myers, and Ye. The students in the R class worked directly with the default R command line interface (Version 3.5.0), and those in the SPSS class were taught to use the SPSS (Version 24) pull-down menus (GUI) for performing all statistical analyses.

The response variable is students' performance, measured by three tests and final exams scores, as well as their average course grades. Once the data collection process was complete, student names and IDs were deleted from the data set to preserve their anonymity.

### 3.3. PROCEDURE AND INSTRUMENTS

On the first day of class, an assessment was administered to both classes to test the students' initial level of mastery of math skills necessary for the course. Students were asked to solve a linear equation, to simplify expressions containing factorials and summations, to integrate a monomial, to answer word problems, to create a linear graph, to interpret a line graph in the context of a word problem, and to answer questions regarding any past experience with statistics or statistics software as well as the most recent math course they completed. Students were also notified that the assessment would have no impact on their grade and that their scores will remain confidential.

The same instructor (the principal investigator of the study), with similar proficiency in both packages, taught both classes and covered the same course material with equivalent class-time allocation. The three exams and final were given to both classes at weeks 4, 8, 12, and 16, respectively. In addition, all exams were 100-point tests, and were similar in form and content (but not identical) for the two classes as they were created to evaluate the same knowledge base.

The course was taught under the lecture format with technology-assisted instruction, and a computer lab was reserved for the entire semester as the regularly scheduled classroom for both classes. In the first week, students learned some commonly used techniques of descriptive statistics, and were then introduced to the software. The instructor first gave the general background of the statistical software (the interface, the cost, the availability, etc.), and then provided a tutorial on using the software, which included how to enter a data set and how to perform basic tasks with the software (such as finding a mean or standard deviation, constructing a frequency distribution, making a bar graph or a histogram, etc.). Step-by-step instructions on performing specific tasks using the software were posted on the course website via the university's learning management system. After the tutorial, students were given a data set containing both qualitative and quantitative variables and asked to explore the data by conducting descriptive statistical analyses for summarizing and visualizing the data set.

During the semester, the instructor integrated the technology into the course in a variety of ways. These included 1) incorporating simulations and game-based learning activities into classroom instruction to help students understand statistical concepts and methods (for example, introducing conditional probability by simulating the famous Monty Hall problem); 2) visualizing abstract concepts using graphical tools (for example, using interactive graphical diagrams in web pages via R "Shiny" apps to improve the teaching of sampling distributions and the Central Limit Theorem); 3) investigating real-life data (of interest and relevance to the students) with the software to help students develop statistical reasoning skills as well as to make learning meaningful. Students were allowed to use software for homework, quizzes, and exams, unless it was specifically stated that hand calculations were required. To provide students with additional practice on the use of technology as well as increase students' motivation to use the software, some of the assigned homework problems were based on raw data rather than summary statistics, and the term project required students to collect data based on the formulation of a research question and conduct appropriate statistical analyses with the software to address the research question.

Toward the end of the semester, a survey (see Appendix) was administered to both classes to assess students' opinions on the use of the statistical software packages. Students were asked how often, if at all, they used the software for graded assignments. They were also asked to use a Likert scale to quantify their feelings about the software (e.g., difficulty or ease of use, impact on the course, and recommendation for the future), as well as the opportunity to provide qualitative feedback. The results provide some information about students' abilities to use the software as well as their attitudes toward the software.

### 3.4. DATA ANALYSIS

Data were summarized and analyzed using *RStudio* (R 3.6.1). Descriptive statistics (mean and standard deviation) were used to summarize results. Independent *t*-tests were performed to compare students' performances on each exam between the two classes. Students' overall semester average in each class was calculated by dividing the total earned points (on quizzes, exams, and term project) by the total possible points in the class, and an independent *t*-test was applied to compare students' semester average between the two classes.

We applied a mixed analysis of variance (ANOVA) to compare the two statistical software packages on students' learning outcomes by testing for differences in exam scores at different testing periods as well as possible testing for an interaction between type of software and testing period. We performed a two-way mixed ANOVA on the data using type of software (R vs. SPSS) as the between-subjects variable and testing periods (assessment, three midterms, and final exam) as the within-subjects variable. Given the noticeable difference in students' performance on Exam 1 between the two groups, a follow-up study was conducted to explore specifically the effect of software on Exam 1 test scores while controlling for the effect of original assessment. For all statistical tests, assumptions were verified and *p*-values less than 0.05 were considered significant.

## 4. RESULTS

### 4.1. SUMMARY STATISTICS

The sample mean test scores (out of a total of 100 points) on the assessment, three exams, and the final exam for both classes are displayed in Table 1. Also included in the table are the overall class average per semester for students in both classes and the results from the independent *t*-tests. As can be seen in the table, students in the two classes exhibit similar level of mathematical skills at the beginning of the semester, as indicated by the similar average scores on the assessment test. Although no statistically significant difference was found between the two groups on tests except for Exam 1, in general, students taught using R scored higher on all three tests and final exam than students taught using SPSS, and showed a better overall performance as indicated by the higher mean semester average.

*Table 1. Comparisons of students' performance by software and testing periods*

| Testing Periods | R Group ($n$ = 30) | SPSS Group ($n$ = 31) | *t*-value (*p*-value) |
|---|---|---|---|
| | Mean (SD) | Mean (SD) | |
| Pre-assessment | 79.17 (16.74) | 78.58 (17.21) | 0.0952 (0.9245) |
| Exam 1 | 89.47 (8.65) | 80.84 (18.22) | 2.3744 (0.0221) |
| Exam 2 | 85.92 (11.13) | 85.26 (14.51) | 0.1993 (0.8428) |
| Exam 3 | 90.48 (10.37) | 88.32 (9.09) | 0.8640 (0.3912) |
| Final exam | 89.85 (8.29) | 88.90 (11.76) | 0.3643 (0.7170) |
| Semester average | 89.41 (7.57) | 87.81 (8.92) | 0.7563 (0.4525) |

Figure 1 shows students' test performance over time and their overall semester performance for both classes. It is clear that students in the two classes had similar initial mathematical ability at the beginning of the semester, and while students in both classes demonstrated improvement on test performance throughout the semester, students in the class using R overall performed slightly better. It is also noted that SPSS students generally exhibited greater variation in test performance compared with the students taught using R.

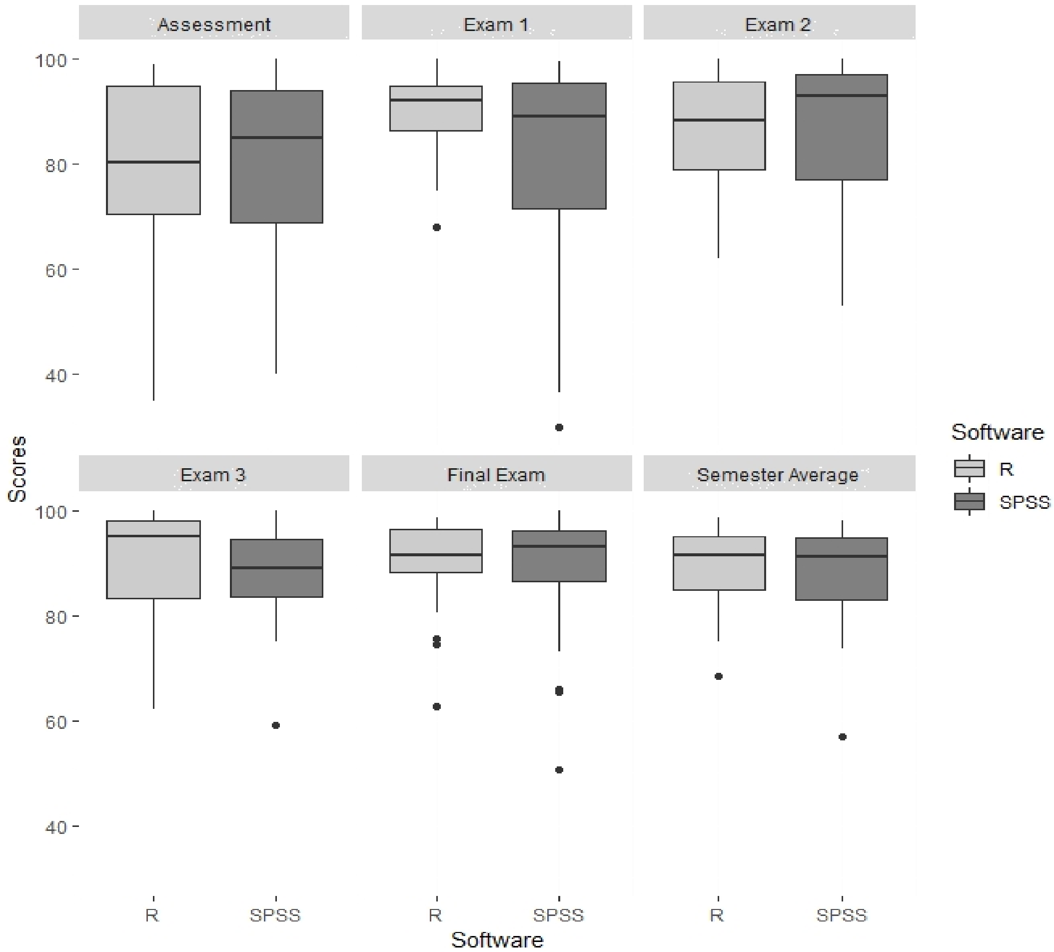


*Figure 1. Students' exam and final performance by software types*

## 4.2. TWO-WAY MIXED ANOVA STUDY

A two-way mixed ANOVA (software × testing periods) was computed on the data to examine the effect on learning outcomes across the two groups and over time. In checking model adequacy, we noticed, after performing Mauchly's sphericity test, that violation of sphericity for the within-subjects factor occurred ($w = 0.0569$, $p < 0.0001$). Thus, we applied Greenhouse-Geisser corrections to the degrees of freedom of the $F$-distribution in ANOVA test to produce more valid test results. The Mauchly's sphericity test was performed using an R function *ezANOVA* provided by the R package *ez*.

The ANOVA yielded a significant main effect of testing periods ($F(5, 295) = 9.83$, $p < 0.0001$ (after using Greenhouse-Geisser corrections)), showing an overall increase in student performance in both groups over time. Although there was a trend for group differences in testing scores (Figure 1), this trend was not significant ($F(1, 5) = 2.84$, $p = 0.1517$), indicating no significant differences in testing scores between the two groups. In addition, no significant software × testing periods interaction was evident ($F(5, 295) = 1.738$, $p = 0.1640$), as indicated in Figure 1 by the comparable response patterns between the two groups at each time point.

Although the overall ANOVA showed no significant difference in testing scores between the two groups, it was interesting to note that students who used R scored significantly higher on Exam 1 than those who used SPSS. To explore the effect of software on students' performance on Exam 1 while controlling for their initial math abilities, we performed a one-way Analysis of Covariance (ANCOVA) with software as the independent variable and assessment scores as the covariate. The ANCOVA results indicated that there was a significant difference in the Exam 1 scores between the two groups while taking into account the influence of the initial assessment ($p = 0.0196$), but there was not a significant interaction between type of software and initial assessment scores ($p = 0.298$). Figure 2 plots students' Exam 1 scores against their original assessment scores for each group, along with the fitted regression lines, and the findings based on the plot are consistent with the ANCOVA results. The positive slopes of the regression lines indicate that in general students with higher scores at initial assessment performed better on Exam 1. The *y*-intercept for the R class is larger than for the SPSS class, indicating that on average students who used R scored higher on Exam 1 than those who used SPSS when adjusted for initial assessment scores. Although not significant, the plot shows some evidence that the effect of software is lower for students with higher assessment scores than it is for those with lower assessment scores. Specifically, students who had better understanding of math/statistics entering the course generally performed well on the first exam, irrespective of the type of software used in class, but for students with weaker math/statistics skills, their performance was dependent more on what software they used when taking the course. It is further noted that students in the R class performed consistently well on Exam 1 while there was considerable variation in student performance within the SPSS group on Exam 1.

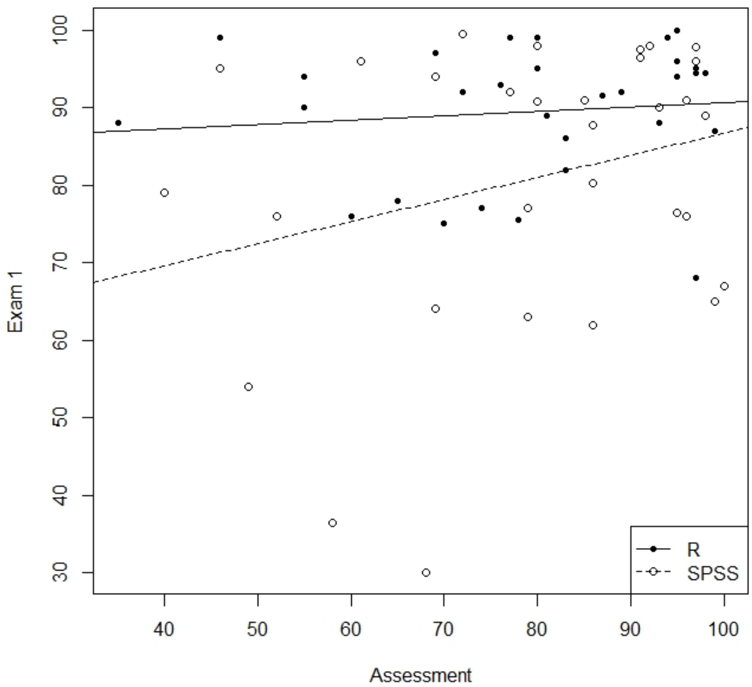


*Figure 2. Plot of exam 1 scores vs. assessment for R and SPSS Groups*

### 4.3. SUMMARY ON THE USE OF SOFTWARE

The end-of-semester survey contained five technology-related questions (four multiple choice questions and one open-ended question). Students were asked how often they used the software on quizzes or exams and whether they used the software on term project, as well as a few questions about their opinions on the software. Students could give multiple responses for the question on the use of software (for example, a student may indicate that he/she used software on three quizzes, two exams, and also on the term project), but we created two mutually exclusive categories for analysis, yes (if a student used the software at least once on graded work) and no (if a student did not use the software at all). To gain insights into students' perception of using software in learning statistics, we asked the students to rate the overall ease of use on a 10-point scale (1 is very difficult and 10 is very easy), and the impact of software on course quality on a 10-point scale (1 is negative and 10 is positive); and we also asked students to choose a number from 1 to 10 to indicate how likely they would recommend the use of the [R/SPSS] software in learning statistics (1 is not at all likely and 10 is very likely). The open-ended question on the survey asked students to offer additional information with regard to the use of the software (comments, concerns, suggestions, etc.), which allowed students to describe their views in their own voice.

Side-by-side boxplots are used to illustrate the differences in students' perceptions regarding the use of software between the two groups (Figure 3). With regard to the perceived ease of use (see boxplots in the bottom left corner of Figure 3), the boxplot for the R group is comparatively short (median = 7.5), which suggests that overall, students had a high level of agreement (most students found R easy to use), with an exception of a few students (shown as outliers on the boxplot). On the other hand, the comparatively tall boxplot for the SPSS group (median = 7) indicates that students held a wide range of opinions regarding the ease of use of SPSS. When asked whether the software had a positive or negative impact on the quality of the course, most students, regardless of group, responded that the use of software improved the quality of the course (see boxplots on the bottom right corner of Figure 3). It is also noted that the boxplot from the R group is slightly right skewed (with a min of 5 out of 10, and median = 7), which suggests that a few students who used R felt particularly strongly that the use of R had improved the course quality. On the contrary, a left skewed distribution is observed in the boxplot for the SPSS group (median = 8), indicating most students agreed that the use of SPSS had a positive impact on the course quality while some students held different opinions (including a 4 and a 1). In comparing students' responses to the question of whether they would recommend the software for future use, a similar pattern seen in the responses for perceived ease of use is demonstrated in the boxplots in the top left corner of Figure 3; that is, students in the R group shared more similar opinions whereas students who used SPSS seemed to disagree more with each other about whether or not the software should be recommended for future use. It can be seen that both boxplots are left skewed, indicating that most students would likely recommend the use of the software in learning statistics, whereas a few students felt differently. Additionally, students in the R group appeared to be more likely to recommend the future use of the software, as the lower quartile for the R group (median = 9) equals the median for the SPSS group (median = 8) .

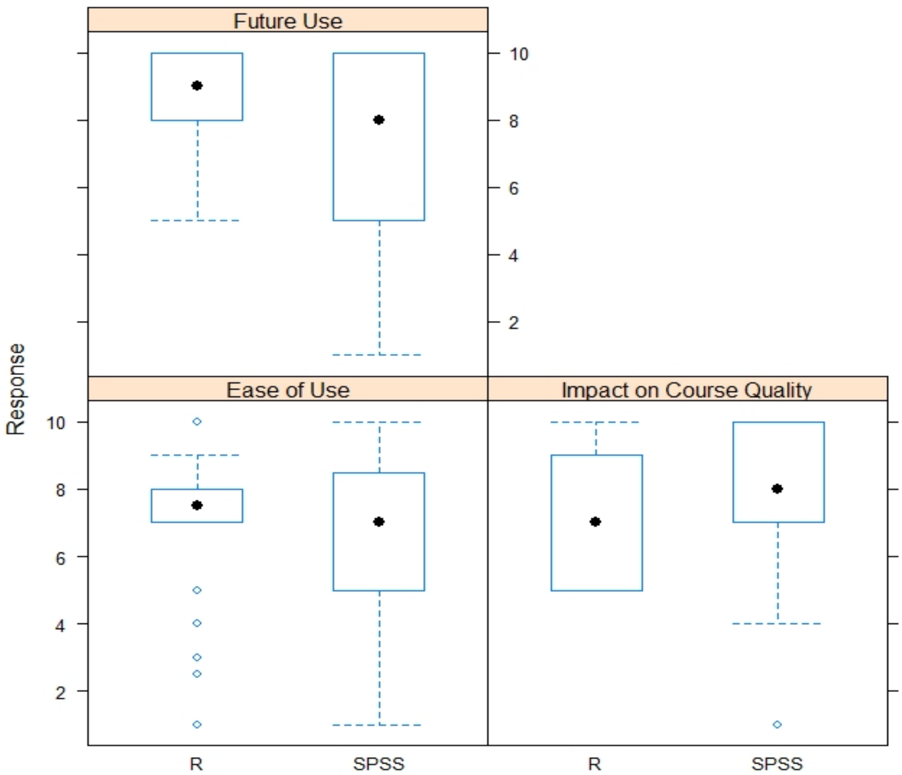


*Figure 3. Comparing students' opinions on the use of software*

The summary statistics of students' responses are presented in Table 2. As shown in the table, the percentages of students who used the software (R or SPSS) when taking quizzes or exams are similar. With respect to the ease of use, the mean ratings from the two groups were moderately high, suggesting that overall students had no difficulty using R or SPSS; furthermore, mean ratings between the two groups did not differ significantly, as the two mean ratings were very close. The survey results indicated that students in both classes viewed using software as having a positive effect on course quality, as evidenced by the high mean ratings of perceived impact. In fact, as indicated by students' responses to the open-ended question, most students thought that the use of software was helpful to their understanding of statistics and to solving problems on a quiz or an exam. As for recommendation for future use, it appeared that, although students in both groups would like to recommend the software for future use, students with experience of using R felt more strongly about recommending the use of R in learning statistics than those in the SPSS class. However, when independent *t*-tests were run to compare students' responses from the two groups, no significant differences were found between the means of the two groups with respect to students' views on ease of use, impact on course quality, or future use of the software ($p > 0.05$).

*Table 2. Summary statistics on software use*

| Item | R Group ($n = 25$) | SPSS Group ($n = 24$) |
|---|---|---|
| Used software on quizzes or exams | 76% yes | 75% yes |
| Mean rating (SD) of ease of use of the software (1 is very difficult and 10 is very easy) | 6.84 (2.32) | 6.74 (2.39) |
| Mean rating (SD) of the software's impact on course quality (1 is negative and 10 is positive) | 7.05 (2.01) | 7.55 (2.39) |
| Mean rating (SD) of recommendation for future use (1 means no and 10 means yes) | 8.50 (1.82) | 7.19 (2.93) |

Considering that Computer Science (CS) students may feel differently about the use of software compared with non-computer science students, we explored further, comparing students' opinions between the CS and non-CS students. We noted that among the 25 students from the R group who responded to the survey, only two students majored in CS, whereas 11 students out of 24 SPSS students who took the survey were CS students. Because of the extremely small sample size of CS students as well as highly unbalanced data in the R group, we focused on the students from the SPSS class only for the purpose of comparing between CS and non-CS students on their views concerning the use of the software. The CS and non-CS students had similar mean ratings with respect to their views on the impact of SPSS on course quality (7.38 vs. 7.65) and on the future use (7.13 vs. 7.23). No significant difference was found between the mean rating on ease of use of SPSS between the CS and non-CS students, although the mean rating for the CS students was a little lower than that for the non-CS students (5.88 vs. 7.26), suggesting that in general non-CS students have found SPSS easier to use compared with the CS students. These findings were not too surprising given the fact that SPSS requires no coding skills and is user-friendly.

## 5. DISCUSSION

The primary goals of this research study were to compare the effectiveness of two statistical software packages (R and SPSS) on students' learning outcomes, and to gather evidence that might support the use of free software in teaching introductory statistics courses, rather than expensive commercial software products like SPSS.

In testing the effectiveness of software, we controlled a number of potential confounding variables. We conducted the study in two different sections of the same statistics course, which were taught by the same instructor. The same course material was covered and similar tests were given in two classes. We performed independent *t*-tests and a mixed ANOVA test to determine the effect of type of software on the response—students' performance measured by their test scores. Our overall findings indicate that using the software package R in teaching introductory statistics is at least as effective as using

SPSS; in fact, the students who used R earned higher grades on all tests compared to those taught using SPSS, even though no statistical significance was found except on Exam 1.

It is noted that the main challenge in teaching statistics with the software package R is that the software requires some programming skills, which can be somewhat intimidating for some students. Because of its command line interface, R requires writing functions and scripts, not just pointing and clicking like SPSS. Some students in the R class told the instructor at the beginning of the semester they were worried about using R because of their lack of experience in language programming and would feel more comfortable using software like SPSS or Excel with pull-down menus. Although it is true that R has a steeper learning curve than SPSS, we found that, with some learning and training in coding, students perceived the same level of difficulty in using R and SPSS, as indicated in the end-of-semester survey. Furthermore, many students who used R responded that they have experienced the strengths of R in data analysis and would highly recommend it for future use.

There are a wealth of resources available (online tutorials, articles, books, etc.) for readers who are interested in learning more about the two software packages and would like to try the software (R and/or SPSS) for teaching introductory statistics. Readers interested in R may refer to the following useful resources: DataCamp (online tutorials, available at https://www.datacamp.com/courses/free-introduction-to-r), DataFlair (online tutorials, available at https://data-flair.training/blogs/r-tutorials-home/), *Beginning R: The Statistical Programming Language* by Mark Gardener (1st edition, 2012), *Learning Statistics Using R* by Randall E. Schumacker (1st edition, 2014), and *Introduction to Scientific Programming and Simulation Using R* by Owen Jones, Robert Maillardet, and Andrew Robinson (2nd edition, 2014). For readers interested in SPSS, here are some great resources: SPSS Tutorials (online tutorials, available at https://www.spss-tutorials.com/basics/), *A Guide to SPSS, Data Analysis and Critical Appraisal* by Belinda Barton and Jennifer Peat (2nd edition, 2014), *Discovering Statistics Using IBM Statistics* by Andy Field (5th edition, 2017), and *Using IBM SPSS Statistics: An Interactive Hands-On Approach* by James Aldrich (3rd edition, 2018).

## 5.1. LIMITATIONS

Due to the limited ability to gain access to students enrolled in an introductory statistics course at the university, the present study included a sample of students enrolled in the statistics course the principal investigator taught in Fall 2017 and Spring 2018. Therefore, the sampling technique employed in this study was convenience sampling rather than random sampling, and a potential problem that came with such convenience was that the students who participated in the study may not truly be a random sample. Particularly, we noted that all participants were students with a STEM background—a group of students with good mathematics and science skills—and hence were not representative of students in most introductory statistics courses who often do not possess high level skills in mathematics, science, and technology. Therefore, one avenue for further research would be to broaden the study to include students who have more diverse backgrounds and hence more accurately represent the variety of students commonly present in most introductory statistics courses. In addition, the scope of the present study was limited to the impact of two software packages in an introductory statistics course taught in two different terms. There are a variety of other college-level statistics courses besides the introductory statistics (e.g., mathematical statistics, intermediate statistics, advanced statistics), and the effect of technology may vary among these different courses. For a more comprehensive understanding of how different technology impacts students' learning outcomes, further work is needed to investigate the effectiveness of software programs in different types of statistics courses.

## 5.2. CONCLUSIONS

The small scale research study in this article, far from being a full and complete guide for evaluating the two statistical software packages (R and SPSS), was intended to provide some insights on the use of different software packages in statistics education, particularly on the learning outcomes and user-friendliness of statistical software. More detailed and large scale research in this area is needed, and we hope that the findings presented here will provide evidence that might encourage further investigation in using free software for learning and teaching statistics.

**REFERENCES**


Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. In M. A. Clements, A. J. Bishop, C. Keitel, J. Kilpatrick, & F. K. S. Leung (Eds.), *Third international handbook of mathematics education* (pp. 643–689). New York: Springer.

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education Journal*, *1*(1). [Online: http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art2/]

Chen, A. C., Moran, S., Sun, Y., & Vu, K. L. (2018). Comparison of Intellectus Statistics and Statistical Package for the Social Sciences. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics*. EPCE 2018. Lecture Notes in Computer Science, vol 10906. New York: Springer

Fox, J., & Bouchet-Valat, M. (2018). Rcmdr: R Commander. R package version 2.5-1. [Online: http://repositories.cdlib.org/uclastat/cts/tise/vol1/iss1/art2]

Fox, J. (2017). *Using the R Commander: A point-and-click interface for R*. Boca Raton, FL: Chapman and Hall/CRC press.

Fox, J. (2005). The R Commander: A basic statistics graphical user interface to R. *Journal of Statistical Software*, *14*(9), 1–42.

Franklin, C., & Garfield, J. (2006). The GAISE (Guidelines for Assessment and Instruction in Statistics Education) project: Developing statistics education guidelines for pre K-12 and college courses. In G. Burrill (Ed.), *2006 NCTM Yearbook: Thinking and reasoning with data and chance*. Reston, VA: National Council of Teachers of Mathematics.

GAISE College Report ASA Revision Committee. (2016). *Guidelines for assessment and instruction in statistics education college report 2016*. Alexandria, VA: American Statistical Association. [Online: http://www.amstat.org/education/gaise]

Garfield, J. B. (1995). How students learn statistics. *International Statistical Review*, *63*(1), 25–34.

Gomez, R. (2010). Innovations in teaching undergraduate statistics courses for biology students. *Review of Higher Education and Self-Learning, 3*(7), 8–13.

Gomez, R. (2011). Teaching a second statistics course for undergraduate psychology students. *Review of Higher Education and Self-Learning, 4*(11), 1–6.

Higazi, S. M. (2002). Teaching statistics using technology. *Proceedings of Journal of Criminal Justice Education, 13*, 433–442. [Online: https://iase-web.org/documents/papers/icots6/5f4_soha.pdf]

IBM SPSS Statistics for Windows, Version 25.0 (Released 2017). IBM Corp., Armonk, NY: IBM Corp. [Online: https://www.ibm.com/analytics/spss-statistics-software]

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics, 5*(3), 299–314.

Koh, K., & Witarsa, M. (2003). A review of data analysis for the behavioral sciences using SPSS. *Journal of Educational and Behavioral Statistics*, *28*(1), 83–88.

McNamara, A. A. (2015). *Bridging the gap between tools for learning and for doing statistics* (Unpublished doctoral dissertation). University of California – Los Angeles, USA. [Online: http://www.amelia.mn/dissertation.html]

Muenchen, R. A. (2014). *The popularity of data analysis software*. [Online: http://r4stats.com/articles/popularity/]

Nunes, C. A., Alvarenga, V. O., Sant'ana, A. S., Santos, J. S., & Granato, D. (2015). The use of statistical software in food science and technology: Advantages, limitations, and misuses. *Food Research International*, *75*, 270–280.

Ozgur, C., Kleckner, M., & Li, Y. (2015). Selection of statistical software for solving big data problems: A guide for businesses, students, and universities. *SAGE Open*, 1–12.

Ozgur, C., Dou, M., Li Y., & Rogers, G. (2017). Selection of statistical software for data scientists and teachers. *Journal of Modern Applied Statistical Methods, 16*(1), 753–774.

Paura, L., & Arhipova, I. (2012). Advantages and disadvantages of professional and free software for teaching statistics. *Information Technology and Management Science, 15*(1), 9–14.

Proctor, J. (2002). SPSS vs. Excel: Computing software, criminal justice students, and statistics. *Journal of Criminal Justice Education*, *13*(2), 433–442.

R Core Team (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.
[Online: www.R-project.org]

R Foundation (2018). *R: Regulatory Compliance and Validation Issues A Guidance Document for the Use of R in Regulated ClinicalTrial Environments* 5.1.1, 5.1.2.
[Online: www.r-project.org/doc/R-FDA.pdf]

RStudio Team (2015). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA.
[Online: http://www.rstudio.com]

Valero-Mora, P. M., & Ledesma, R. (2012). Graphical user interfaces for R. *Journal of Statistical Software*, *49*(1).
[Online: https://www.jstatsoft.org/index.php/jss/article/view/v049i01/v49i01.pdf]

Wang, C., Chen, M. H., Schifano, E., Wu, J., & Yan, J. (2016). Statistical methods and computing for big data. *Statistics and its Interface*, *9*(4), 399–414.

Wilhelm, A. (2007). Use R for teaching statistics in the social sciences?! *Proceedings of the 56th session of the International Statistical Institute*, Lisbon, Portugal. Voorburg, The Netherlands: International Statistical Institute.
[Online: https://iase-web.org/documents/papers/isi56/CPM81_Wilhelm.pdf]

LUCY KERNS
Youngstown State University
One University Plaza
Youngstown, OH 44555

**APPENDIX**

Name:____________________________________________
Date:____________________

**Instructions:** Please fill out the survey below. This is not your instructor assessment, but rather a survey to gather information on the proceedings of the semester, most importantly the use of the computer program. Note that this assessment will **NOT** affect your grade for the semester in any way. Please do not refrain from giving any and all negative and/or positive feedback relevant to each question.

1. How often did you use R throughout the semester? (circle and complete all that apply)
   a. I used R on _____ quizzes.
   b. I used R on _____ exams.
   c. I used R for my term project.
   d. I did not use R at all.

2. How easy or difficult was it to successfully use R throughout the semester? Please rate, on a scale from 1 to 10, where a 1 is considered to be very difficult while a 10 is considered to be very easy, the level of difficulty you experienced with R this semester and explain why on the lines below. Please write “N/A” if you did not use R this semester.

   ____________________________________________________________

   ____________________________________________________________

3. Do you feel that R improved or detracted from the quality of the course? Please rate, on a scale from 1 to 10, where a 1 is considered to mean that R interfered with the quality of the course while a 10 is considered to mean that R greatly added to the quality of the course, the impact R had on your experience this semester and explain why on the lines below. Please write “N/A” if you did not use R this semester.

   ____________________________________________________________

   ____________________________________________________________

4. Would you recommend the further use of R in this course moving forward? Please rate, on a scale from 1 to 10, where a 1 is considered to be a definite no while a 10 is considered to be a definite yes, your recommendation for the use of R for this course moving forward and explain why on the lines below. Please write “N/A” if you did not use R this semester.

   ____________________________________________________________

   ____________________________________________________________

5. Please use the lines below to make any comments or to express any concerns about the use of R in this course that were not addressed in the previous questions.

   ____________________________________________________________

   ____________________________________________________________

   ____________________________________________________________