# Introduction

## Rationale

The practice of statistics is computational. Statistical computing software are mainly used to do statistical computing. With the emergence of data science as a field due to big data, machine learning and powerful computers that are no longer expensive all brought about by the 4th industrial revolution, it is very fitting to consider how statistics education should adapt to these changes and be relevant in the practice of statistics. In this regard, in 2005, Franklin et al. (2007) put forth the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report recommending a framework for statistics education both in the k to 12 and college level. In 2016, the GAISE College Report ASA Revision Committee (2016) revisited the effectiveness of the framework and still found it effective. The framework is now the standard in statistics education in the United States and in many countries that adopted it (Zeiffler, Garfield, & Fry, 2018).

In the GAISE report, one noteworthy recommendation is “the use of technology to explore concepts and analyze data.” Studies have shown that the use of technology can really improve statistics education; equipping learners with relevant data skills and effective powerful tools in this era where data is very much abundant (Chance, Ben-Zvi, Garfield, & Medina, 2007; Chance & Rossman, 2006; Çetinkaya-Rundel & Rundel, 2017; Doi, Potter, Wong, Alcaraz, & Chi, 2016; Harraway, 2012; Stander & Dalla Valle, 2017). However, when it comes to software used in doing statistical computing and teaching statistics, there is no single statistical computing tool that fits all statistical tasks (McNamara, 2018). Nonetheless, introductory statistics students should be taught a common statistical computing software such as SAS, SPSS, or R (R Core Team, 2018), enthusing them to continuously learn statistics technology since statistical tools are diverse and eventually evolve through time (Gould et al., 2018). Moreover, Gould (2010), N. J. Horton et al. (2015), N. J. Horton (2015) and Hardin et al. (2015) pointed out the importance of developing among students – with the use of statistical computing software – data management skills in introductory and second courses in statistics.

In our country, leading universities are addressing this issue and have already integrated the use of technology in their curriculum for statistics education. The University of the Philippines for example uses a number of software in its introductory statistics courses and statistics courses (eg. R with RStudio, Python, SAS, SPSS, Stata, MS Excel, QGIS, ArcGIS, Gephi, yEd Graph Editor, and more). Most of the software used are opensource.

In the province, some universities have acquired SPSS to teach statistics courses. On the other hand, some still uses calculators to do and teach statistical computing. Unfortunately, both setup do not lessen the gap between statistics education and statistical practice. In reality, most institutions to which graduates from these universities get employed cannot afford SPSS. One can choose MS Excel as an alternative, however its functions are limited (Biehler, Ben-Zvi, Bakker, & Makar, 2013).

These developments gave way to the emergence of data science as a field. As a result, the practice of statistics has dramatically changed and has distanced away from statistics education (Finzer, 2013; Wood, Mocko, Everson, Horton, & Velleman, 2018; Zeiffler et al., 2018). Nonetheless, some measures are already in place to lessen the gap between statistical practice and statistics education.

Moreover, Gould (2010), Horton et al. (2015), N. J. Horton (2015) and Hardin et al. (2015) pointed out the importance of developing among students data management skills in introductory and second courses in statistics.

Reproduciblequite

The gap between statistical practice and statistics education.

barriers in using technology (price)

R and RStudio

Efforts in the Philippines

## Statement of the Problem

1. What are the profile variables of the respondents?
2. What are the calculator test scores and RStudio test scores of the respondents?
3. What are the calculator test scores and RStudio test scores of the respondents when grouped according to the profile variables?
4. Is there significant difference in the calculator test scores and the RStudio test scores of the respondents when grouped according to the profile variables?
5. Is there significant relationship between age and calculator test scores and age and RStudio test scores of the respondents?
6. Is there significant difference between the calculator test scores and RStudio test scores of the respondents?

## Statement of the Hypothesis

1. Is there significant difference between the calculator test scores and RStudio test scores of the respondents?
2. the

## Significance of the Study

## Research Framework

## Scope

## Definition of Terms

# Review of Related Literature and Studies

The choice of which software to use in teaching statistics can be quite a challenge given a lot of things to consider. It can be a problem similar to bridging the gap between the practice of statistics and statistics education. Tools for learning statistics can be used easily by starters opposite to when using tools for doing statistics (Gould et al., 2018). However,

# Methodology

## Research Design

## Research Environment

## Respondents

## Research Instruments

## Data Gathering Procedure

## Data Analysis

# Results and Discussion

## Profile Variables of the Respondents

The profile variables of the respondents considered in the study are gender: male or female, have previous programming experience: yes or no, statistical computing preference: calculator or RStudio, and age.

### Gender

Most of those who enroll BSEd mathematics are female. It is no surprise that among the respondents, 62.50% are female and 37.50% are male. Table 1 shows the frequency distribution of gender.

**Table 1:** Frequency Distribution of Gender

|  |  |  |
| --- | --- | --- |
| Gender | Freq | Percentage |
| Female | 5 | 62.5 |
| Male | 3 | 37.5 |
| Total | 8 | 100 |

### Have Previous Programming Experience

Among the respondents, half have previos programming experience while the other half have none. Table 2 shows the frequency distribution for have previous programming experience. It seems that some students still have no programming experience when they take their second course in statistics.

**Table 2:** Frequency Distribution of Have Previous Programming Experience

|  |  |  |
| --- | --- | --- |
| Experienced | Freq | Percentage |
| No | 4 | 50 |
| Yes | 4 | 50 |
| Total | 8 | 100 |

### Statistical Computing Tool Preference: Calculator or RStudio

Having learned the basics of using calculator and RStudio for statistical computing and took the examination using calculator and RStudio and then knowing their scores, half of the respondents still preferred to use calculator while the other half now preferred to use RStudio. Table 3 shows the frequency distribution for statistical computing tool preference. Among those who preferred RStudio, one said “it is easy”. Among those who preferred calculator, one found using RStudio “complicated” and that using calculator is “comfortable”. Table 3 shows the frequency distribution of statistical computing tool preference.

**Table 3:** Frequency Distribution of Statistical Computing Tool Preference

|  |  |  |
| --- | --- | --- |
| Preference | Freq | Percentage |
| Calculator | 4 | 50 |
| RStudio | 4 | 50 |
| Total | 8 | 100 |

### Age

Half of the respondents are 21 years old, two are 22 years old, one is 24 years old and one is 39 years old. Table 4 is the frequency distribution of age.

**Table 4:** Frequency Distribution of Age

|  |  |  |
| --- | --- | --- |
| Age | Freq | Percentage |
| 21 | 4 | 50 |
| 22 | 2 | 25 |
| 24 | 1 | 12.5 |
| 39 | 1 | 12.5 |
| Total | 8 | 100 |

## Calculator Test Scores and RStudio Test Scores of the Respondents

There are two sets of examination scores: the calculator test scores and RStudio test scores.

### Calculator Test Scores

The result of the calculator test shows that 62.50% failed and 37.50% passed. The mean score is 46.25% which is not even passing. The exam result is not remarkable. Table 5 shows the frequency distribution and summary of calculator test scores.

**Table 5:** Frequency Distribution and Summary of Calculator Test Scores

|  |  |  |
| --- | --- | --- |
| Scores | Freq | Percentage |
| Failed | 5 | 62.5 |
| Passed | 3 | 37.5 |
| Total | 8 | 100 |

|  |  |  |  |
| --- | --- | --- | --- |
| Min | Mean | Max | SD |
| 25 | 46.25 | 100 | 25.06 |

### RStudio Test Scores

The result of the RStudio test shows that 37.50% passed and 62.50% failed. The mean score is 54.50 which is passing. Compared to the calculator test scores, the RStudio test scores seem to have a better result. Table 6 shows the frequency distribution and summary of RStudio test scores.

**Table 6:** Frequency Distribution and Summary of RStudio Test Scores

|  |  |  |
| --- | --- | --- |
| Scores | Freq | Percentage |
| Failed | 3 | 37.5 |
| Passed | 5 | 62.5 |
| Total | 8 | 100 |

|  |  |  |  |
| --- | --- | --- | --- |
| Min | Mean | Max | SD |
| 24 | 54.5 | 86 | 21.23 |

## Calculator Test Scores and RStudio Test Scores of the Respondents When Grouped According to the Profile Variables

The calculator test scores and the RStudio test scores of the respondents are grouped according to the profile variables gender, having previous programming experience and statistical computing tool preference.

### Calculator Test Scores According to Gender

Among the female respondents, only one passed the test. Among the male respondents, only one failed the test. The mean score for female respondents is 37 which is not passing. The mean score for male respondents is 61.67 which is passing. Table 7 shows the frequency distribution and summary of calculator test scores according to gender. The calculator test scores of male respondents seems to be better than the scores of female respondents.

**Table 7:** Frequency Distribution and Summary of Calculator Test Scores According to Gender

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Gender | Failed | Passed | Min | Mean | Max | SD |
| Female | 4 | 1 | 25 | 37 | 54 | 11.94 |
| Male | 1 | 2 | 27 | 61.67 | 100 | 36.64 |

### RStudio Test Scores According to Gender

Among the female respondents, the number of respondents who failed the test is one respondent greater than the number of respondents who passed the test. All of the male respondents passed the test. The mean score for female respondents is 43.20 which is not passing. The mean score for male respondents is 73.33 which is passing. Table 8 shows the frequency distribution and summary of RStudio test scores according to gender. The RStudio test scores of male respondents seem to be better than the scores of female respondents.

**Table 8:** Frequency Distribution and Summary of RStudio Test Scores According to Gender

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Gender | Failed | Passed | Min | Mean | Max | SD |
| Female | 3 | 2 | 24 | 43.2 | 68 | 16.87 |
| Male | 0 | 3 | 61 | 73.33 | 86 | 12.5 |

### Calculator Test Scores According to Have Previous Programming Experience

Half of the respondents who don’t have previous programming experience failed the calculator test and the other half passed. Only one respondent who have previous programming experience passed the test. The mean score for those who don’t have previous programming experience is 54.25 which is passing. The mean score for those who have previous programming experience is 38.25 which is not passing. Table 9 shows the frequency distribution and summary of calculator test scores according to have previous programming experience.

**Table 9:** Frequency Distribution and Summary of Calculator Test Scores According to Have Previous Programming Experience

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Experienced | Failed | Passed | Min | Mean | Max | SD |
| No | 2 | 2 | 27 | 54.25 | 100 | 32.5 |
| Yes | 3 | 1 | 25 | 38.25 | 58 | 15.44 |

### RStudio Test Scores According to Have Previous Programming Experience

Half of the respondents who don’t have previous programming experience failed the RStudio test and the other half passed. Only one respondent who have previous programming experience failed the test. The mean score for those who don’t have previous programming experience is 52.50 which is passing. The mean score for those who have previous programming experience is 56.50 which is passing. Table 10 shows the frequency distribution and summary of RStudio test scores according to have previous programming experience.

**Table 10:** Frequency Distribution and Summary of RStudio Test Scores According to Have Previous Programming Experience

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Experienced | Failed | Passed | Min | Mean | Max | SD |
| No | 2 | 2 | 35 | 52.5 | 86 | 23.39 |
| Yes | 1 | 3 | 24 | 56.5 | 73 | 22.22 |

### Calculator Test Scores According to Statistical Computing Tool Preference

In the calculator test, only one passed among those who prefer calculator as statistical computing tool. Among those who preferred RStudio as statistical tool, half passed and half failed in the calculator test. The mean score for those who prefer calculator is 48.75 which is not passing. The mean score for those who prefer RStudio is 43.75 which is also not passing. Table 11 shows the frequency distribution and summary of calculator test scores according to statistical computing tool preference.

**Table 11:** Frequency Distribution and Summary of Calculator Test Scores According to Statistical Computing Tool Preference

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Preference | Failed | Passed | Min | Mean | Max | SD |
| Calculator | 3 | 1 | 25 | 48.75 | 100 | 35.1 |
| RStudio | 2 | 2 | 27 | 43.75 | 58 | 14.71 |

### RStudio Test Scores According to Statistical Computing Tool Preference

Among those who prefer calculator as statistical computing tool, half passed and half failed the RStudio test. Among those who prefer RStudio as statistical computing tool, only one failed. The mean score of those who prefer calculator is 53.25 which is passing. The mean score of those who prefer RStudio is 55.75 which is also passing. Table 12 shows the frequency distribution and summary of RStudio test scores according to statistical computing tool preference.

**Table 12:** Frequency Distribution and Summary of RStudio Test Scores According to Statistical Computing Tool Preference

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Preference | Failed | Passed | Min | Mean | Max | SD |
| Calculator | 2 | 2 | 24 | 53.25 | 86 | 28.74 |
| RStudio | 1 | 3 | 38 | 55.75 | 73 | 14.86 |

## Significant Difference in the Calculator Test Scores and the RStudio Test Scores of the Respondents when Grouped According to the Profile Variables

The calculator test scores and the RStudio test scores are grouped according to gender and have previous programming experience and statistical computing tool preference.

### Calculator Test Scores

For the calculator test scores of the respondents when grouped according to gender, the Wilcoxon rank-sum test gives a p value of 0.29. Since the p value is greater than the significance level of 0.05, the null hypothesis that the mean calculator test score of male is equal to the mean calculator score of female is not rejected. There is no statistical evidence that the mean score of male of 61.67 is greater than the mean score of female of 37. Gender is not a factor among the respondents when it comes to their performance in the calculator test.

For the calculator test scores of the respondents when grouped according to have previous programming experience, the Wilcoxon rank-sum test gives a p value of 0.56. Since the p value is greater than the significance level of 0.05, the null hypothesis that the mean calculator test score of the respondents who have previous programming experience is equal to the mean calculator score of the respondents who have no previous programming experience is not rejected. There is no statistical evidence that the mean score of the respondents who have no previous programming experience of 54.25 is greater than the mean score of the respondents who have previous programming experience of 38.25. Having previous programming experience is not a factor among the respondents when it comes to their performance in the calculator test.

### RStudio Test Scores

For the RStudio test scores of the respondents when grouped according to gender, the Wilcoxon rank-sum test gives a p value of 0.07. Since the p value is greater than the significance level of 0.05, the null hypothesis that the mean RStudio test score of male is equal to the mean RStudio test score of female is not rejected. There is no statistical evidence that the mean score of male of 73.33 is greater than the mean score of female of 43.20. Gender is not a factor among the respondents when it comes to their performance in the RStudio test.

For the RStudio test scores of the respondents when grouped according to have previous programming experience, the Wilcoxon rank-sum test gives a p value of 0.89. Since the p value is greater than the significance level of 0.05, the null hypothesis that the mean RStudio test score of the respondents who have previous programming experience is equal to the mean RStudio test score of the respondents who have no previous programming experience is not rejected. There is no statistical evidence that the mean score of the respondents who have previous programming experience of 56.50 is greater than the mean score of the respondents who have no previous programming experience of 52.50. Having previous programming experience is not a factor among the respondents when it comes to their performance in the RStudio test.

## Significant Relationship Between Age and the Calculator Test Scores and Between Age and the RStudio Test Scores of the Respondents

### Calculator Test Scores and Age

The computed p value and correlation of calculator test scores and age is 0.004 and 0.88 respectively. Table 13 shows that the p value of 0.004 is less than the significance level of 0.05. The null hypothesis that calulator test scores and age are linearly uncorrelated is rejected. At the 5% significance level, the data provide sufficient evidence to conclude that calculator test scores and age are positively linearly correlated. There is a very strong positive correlation. The older the respondent, the better the calculator test score.

**Table 13:** Pearson’s product-moment correlation: Age and Calculator Test

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Test statistic | df | P value | Alternative hypothesis | cor |
| 4.637 | 6 | 0.003552 \* \* | two.sided | 0.8842 |

### RStudio Test Scores and Age

The computed p value and correlation of RStudio test scores and age is 0.12 and 0.59 respectively. The p value of 0.12 is greater than the significance level of 0.05. The null hypothesis that RStudio test scores and age are linearly uncorrelated is not rejected. At the 5% significance level, the data provide no evidence to conclude that RStudio test scores and age are positively linearly correlated. In their RStudio test scores, age does not matter among the respondents.

## Significant Difference Between the Calculator Test Scores and Rstudio Test Scores of The Respondents

Comparing the mean calculator test score and the mean RStudio test score, the Wilcoxon signed rank-sum test gives a p value of 0.25. Since the p value is greater than the significance level of 0.05, the null hypothesis that the mean calculator test score is equal to the mean RStudio test score is not rejected. There is no statistical evidence that the mean RStudio test score of 54.50 is greater than the mean calculator test score of 46.25. The use of RStudio as statistical computing tool in the test has not significantly improved the performance of the respondents.

# Reference

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. 643–690). 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*, *1*(1). Retrieved from <https://escholarship.org/uc/item/8sd2t4rr>

Chance, B., & Rossman, A. (2006). Using Simulation to Teach and Learn Statistics. In A. Rossman & B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.

Çetinkaya-Rundel, M., & Rundel, C. (2017). Infrastructure and Tools for Teaching Computing Throughout the Statistical Curriculum. *The American Statistician*. <https://doi.org/10.1080/00031305.2017.1397549>

Doi, J., Potter, G., Wong, J., Alcaraz, I., & Chi, P. (2016). Web Application Teaching Tools for Statistics Using R and Shiny. *Technology Innovations in Statistics Education*, *9*(1). Retrieved from <https://escholarship.org/uc/item/00d4q8cp>

Finzer, W. (2013). The Data Science Education Dilemma. *Technology Innovations in Statistics Education*, *7*(2). Retrieved from <https://escholarship.org/uc/item/7gv0q9dc>

Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*. Alexandria, VA: American Statistical Association.

GAISE College Report ASA Revision Committee. (2016). Guidelines for Assessment and Instruction in Statistics Education College Report 2016. Retrieved from <http://www.amstat.org/education/gaise>

Gould, R. (2010). Statistics and the Modern Student. *International Statistical Review*, *72*(2), 297–315. <https://doi.org/10.1111/j.1751-5823.2010.00117.x>

Gould, R., Wild, C. J., Baglin, J., McNamara, A., Ridgway, J., & McConway, K. (2018). Revolutions in Teaching and Learning Statistics: A Collection of Reflections. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 457–472). Cham: Springer.

Hardin, J., Hoerl, R., Horton, N. J., Nolan, D., B. Baumer, Hall-Holt, O., … Ward, M. D. (2015). Data Science in Statistics Curricula: Preparing Students to “Think with Data”. *The American Statistician*. <https://doi.org/10.1080/00031305.2015.1077729>

Harraway, J. A. (2012). Learning Statistics Using Motivational Videos, Real Data and Free Software. *Technology Innovations in Statistics Education*, *6*(1). Retrieved from <https://escholarship.org/uc/item/1fn7k2x3>

Horton, N. J. (2015). Challenges and Opportunities for Statistics and Statistical Education: Looking Back, Looking Forward. *The American Statistician*, *69*(2), 138–145. <https://doi.org/10.1080/00031305.2015.1032435>

Horton, N. J., Baumer, B. S., & Wickham, H. (2015). Taking a Chance in the Classroom: Setting the Stage for Data Science: Integration of Data Management Skills in Introductory and Second Courses in Statistics. *CHANCE*, *28*(2), 40–50. <https://doi.org/10.1080/09332480.2015.1042739>

McNamara, A. (2018). Key Attributes of a Modern Statistical Computing Tool. *The American Statistician*. <https://doi.org/10.1080/00031305.2018.1482784>

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Stander, J., & Dalla Valle, L. (2017). On Enthusing Students About Big Data and Social Media Visualization and Analysis Using R, RStudio, and RMarkdown. *Journal of Statistics Education*, *25*(2), 60–67. <https://doi.org/10.1080/10691898.2017.1322474>

Wood, B. L., Mocko, M., Everson, M., Horton, N. J., & Velleman, P. (2018). Updated Guidelines, Updated Curriculum: The GAISE College Report and Introductory Statistics for the Modern Student. *CHANCE*, *31*(2), 53–59. <https://doi.org/10.1080/09332480.2018.1467642>

Zeiffler, A., Garfield, J., & Fry, E. (2018). What is Statistics Education? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 37–70). Cham: Springer.