

2019

Comparative Analysis of Students' Performance Between Online and on Campus in an Introductory Statistics Course

Kendal McDonald

Georgia College and State University

Follow this and additional works at: <https://kb.gcsu.edu/thecorinthian>

 Part of the [Applied Statistics Commons](#), and the [Categorical Data Analysis Commons](#)

Recommended Citation

McDonald, Kendal (2019) "Comparative Analysis of Students' Performance Between Online and on Campus in an Introductory Statistics Course," *The Corinthian*: Vol. 19 , Article 8.

Available at: <https://kb.gcsu.edu/thecorinthian/vol19/iss1/8>

This Article is brought to you for free and open access by Knowledge Box. It has been accepted for inclusion in The Corinthian by an authorized editor of Knowledge Box.

Comparative Analysis of Students' Performance Between Online and on Campus in an Introductory Statistics Course

Kendal McDonald

Dr. Jebessa Mijena

Faculty Mentor

INTRODUCTION

Comparing on-campus and online student performance in an introductory statistics course gives insight on which class would be better for a student to take given the performance of past students. The motivation behind this research is to see if there is a difference between students' performance between online and on-campus in an introductory statistics and probability course at Georgia College. A student's performance was based on homework and quizzes the student does in MyStatLab for both online and on-campus. The online data was comprised of five sections which were from the semesters of summer 2014, summer 2015, summer 2016, and summer 2017 with summer 2017 having two sections during a semester. For on-campus data, the data was comprised of nine sections from the semesters of spring 2014, spring 2016, and spring 2017 with each semester having three sections. The number of attempts given for each homework and quiz is respectively ten and five. In this paper, the word on-campus and in-class represent the same students and will be used interchangeably. The significance level in the current paper is 0.05.

METHODOLOGY

Through MyStatLab, both on-campus and online sections were able to take all quizzes and homework assignments online at their own convenience. For the online group consisted of five sections from the summer semester of the years 2014, 2015, 2016, and 2017. Each of the years had one section except for 2017 which had two sections. Each summer course was taken over eight conservative weeks. Summer 2014 section had eighteen students who took this class with no students withdrawing from the class; however, for homework and quizzes one student did not attempt majority of the

assignments which was ignored in the data. From the excel sheet downloaded from MyStatLab, line 14 will need to be deleted from both the homework and quiz excel sheets. The summer 2015 section had twenty-two students with no one dropping the course, but one student did not complete a sufficient amount of the homework assignments or the quizzes. Therefore, to clean the data from summer 2015 the deletion of line 8 on the homework and quiz excel sheets were done. Summer 2016 section had twenty-four students with one student withdrawing from the course before the end of the eight-week course was over. Thus, the withdrawing student was the only student to not complete majority of the work which was line 10 on the homework and quiz from the excel sheets generated by MyStatLab which was deleted. For summer 2017, there were two sections the first section, labeled W04, was taken over the course of May and June while the second section, labeled W07, was taken over the course of June and July. The first section of summer 2017 had twenty-four students and the second section had twenty-six students. Between both sections of summer 2017 all students completed majority of the homework and quiz assignments that no data clean up in these sections were needed.

On-campus data was comprised of nine sections from spring 2014, spring 2016, and spring 2017. Each semester had three sections with all sections covering a span of sixteen weeks. When cleaning the data, each data starts on line 2 due to the headings of the columns in line one. For spring 2014, the first and second section, labeled section 11 and section 12 respectively, each had thirty-five students and the third section, labeled section 13, had thirty-three students. To clean the data for spring 2014 section 11, the deletion of lines 19, 29, and 32 are necessary for both homework and quiz excel sheets. Section 12 of spring 2014, for data cleaning the deletion of lines 2, 6, 26, 31, and 32 from the excel sheets for homework and quizzes. For section 13 of spring 2014, removing lines 6 and 24 will clean up the data since these two students either withdrew from the class or did not complete majority of the work after a certain period in the course. For spring 2016, all three sections, labeled 2, 3, and 4 respectively, had thirty-seven students in the course. Section 2 of the semester spring 2016, only two students did not complete sufficient amount of the course work which would be students on lines 8 and 37 thus these lines should be deleted. For section 3 for the semester spring 2016, lines 17 and 21 need to be deleted to clean the data since these students did not finish majority of the course work.

For the last section, section 4, of spring 2016 there was only one student who did not complete the majority of the assignments which is student signified by line 2 was deleted for data cleaning. The next semester the on-campus data is comprised of is spring 2017 with first and third sections, labeled 2 and 13 respectively, have thirty students and the third section, labeled 3, has thirty-one students. To clean the data for section 2 of spring 2017 the removal of line 26 on the excel sheets for both homework and quizzes since the student did not complete a sufficient amount of the course work. For spring 2017 section 3, the removal of lines 20 and 24 are needed for cleaning of the data. For the last section, section 13, of spring 2017, to clean the data the removal of lines 3 and 10.

With these data clean ups completed, the next step in getting the data information needed is by creating columns for the quizzes in the order of days before due date, number of attempts, highest score on the quiz, and the attempt the highest score was achieved for each quiz. After the final quiz and final homework highest score attempt column, the addition of first attempt quiz average and final quiz average should be added for each quiz data set and for homework only final homework average should be added for each quiz data set. The first attempt quiz average is comprised of all the first attempt quiz score a student makes on the quiz then minus the two lowest grades made on a quiz divided by four for online sections and seven for on campus sections except for the semester of spring 2016 which will be six instead of seven. The final quiz average is calculated the same way but instead of the first attempt scores being used the highest score on the quiz is used in the same method. The homework final average is calculated the same way with the two lowest homework grades are dropped then divided by seven except for spring 2016 semester for on-campus and also by seven for online.

The next column beside the days before due date is the number of attempts. The number of attempts column is calculated by looking at the total number of attempts a student takes a certain quiz. If a student did not attempt the quiz, the student does not receive a zero and the box in the column should just remain blank. Also, if a number is negative which means the assignment was completed after due date either delete it or remember to ignore the negatives when completing the analysis.

The following column after number of attempts is the highest score. The highest score is when the student made the highest grade on the quiz. With the highest score

column, all is needed is inputting the highest score the student made. Once again, if a student did not attempt the quiz no grade would be given and the box would be left blank in the column.

Next column is the highest score attempt. This column signifies the attempt the student achieved the highest score on the quiz. If a student made a hundred on their fourth attempt, the number recorded in the column would be four since it was on the fourth attempt the student made their highest score. Also, a student who does not attempt a quiz shall not receive a number in the box thus the box would remain empty.

For finding the first attempt quiz average, final quiz average and homework average, if a student did not attempt the quiz a zero must be given. The best way to find the averages would be copying the columns first score attempt, the first column under the quiz number, and the highest score columns then pasting them in a new excel sheet. Once the data is put into the new excel sheet, for every blank put a zero. The reason behind putting the zeros in is because if there were no zeros then an error would occur in the calculation of the data. Once all zeros are in place, calculating the first attempt quiz average is done by filling in the column with the equation $((\text{sum}(\text{all the columns relating to first attempt quiz average}) - \text{small}(\text{same columns}, 1) - \text{small}(\text{same columns}, 2)) / 7)$ for on-campus course except for the spring 2016 semesters which would be divide by six and for online semesters would be divided by four. The same equation layout will be used for the final quiz average which is based off of the highest score a student makes on a quiz except the columns for the sum and small parts of the equation would now be with the highest scores columns. The homework average is calculated the same way as either quiz grades however for online it will be divided by seven instead of four and spring 2016 will still be divided by six. Once all the data is filled in, all sections data will be combined for both online and on-campus with quiz and homework being in the same excel sheet.

The analysis was conducted in R version 3.3.2 and RStudio version 1.0.316, a statistical programming software. Through Georgia College Institutional Review Board, the approval for gaining and analyzing the data was granted on May 17, 2017.

RESULTS

The Wilcoxon Rank Sum Test is used due to the outliers in the data while also due to the lack of normality in the data. The Wilcoxon Rank Sum Test formulas can be seen in Figure 1 where W is Wilcoxon statistic for the sum of the n signed ranks, μ_W is the mean of W , $\text{Var}(W)$ is the variance of W , n_1 and n_2 are the sample sizes of the two samples:

$$\mu_W = \frac{n_2(n_1 + n_2 + 1)}{2}$$

$$\text{Var}(W) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

$$Z = \frac{W - \frac{n_2(n_1 + n_2 + 1)}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

Figure 1: Wilcoxon Rank Sum Test Formulas

The Wilcoxon rank sum test is a nonparametric alternative to the two-independent sample t-test which is based solely on the order in which the observations from the two samples fall. This applies to any distribution, normal or otherwise. For large samples ($n_1 > 10, n_2 > 10$) Z is approximately standard normal distribution. The alternative hypothesis for Wilcoxon rank sum is the distributions of populations I and II have different locations.

The first comparison analyzed is the final quiz scores. Table 1 gives summaries of quiz average scores. In Figure 2, we can see the final quiz averages are very close in average with on-campus having a higher final average by 0.33%.

Table 1: Summary of Final Quiz Average Scores

Final Quiz Average Online	Final Quiz Average on-Campus
---------------------------	------------------------------

Sample Size: 110	Sample Size: 284
Minimum: 29.77	Minimum: 38.18
1 st Quartile: 85.11	1 st Quartile: 82.03
Median: 89.47	Median: 90.42
Mean: 87.02	Mean: 87.35
3 rd Quartile: 94.37	3 rd Quartile: 95.14
Maximum: 100.00	Maximum: 100.00

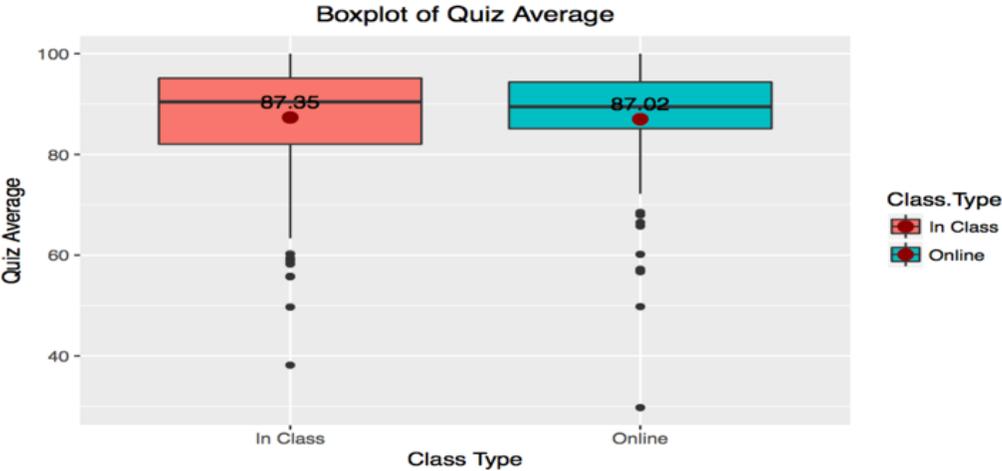


Figure 2: Boxplot of Final Quiz Average

By the Wilcoxon Rank Sum Test, the calculated p-value is 0.7615 which is greater than 0.05 therefore the distributions of final quiz average scores for online and on-campus are not significantly different.

Similar to the final quiz average, the first attempt quiz average behaved very similar as shown in Table 2 and Figure 3. The p-value calculated for the first attempt quiz average was 0.8280 which shows online and on-campus are not significantly different in distribution.

Table 2: Summary of First Attempt Quiz Average

1 st Attempt Quiz Average Online	1 st Attempt Quiz Average on-Campus
Sample Size: 110	Sample Size: 284
Minimum: 23.12	Minimum: 12.87
1 st Quartile: 57.03	1 st Quartile: 54.83
Median: 67.84	Median: 68.28
Mean: 65.08	Mean: 64.21
3 rd Quartile: 76.93	3 rd Quartile: 76.00
Maximum: 93.42	Maximum: 97.13

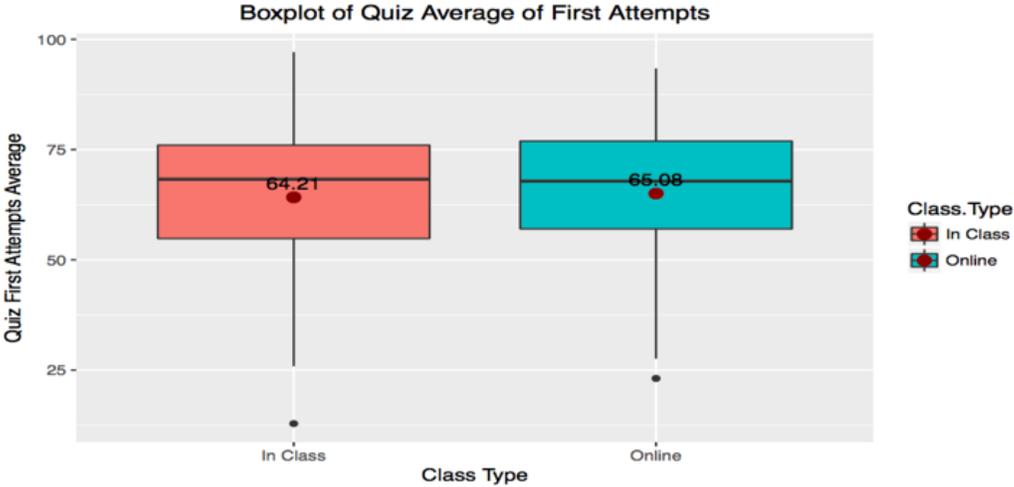


Figure 3: Boxplot for First Attempt Quiz

Differently from the first and final quiz average, the homework averages vary by a difference of 3.19 % with on-campus having the higher average as shown in Table 3. Looking at Figure 4, the boxplot shows there are many outliers in both online and on-campus. These outliers would cause a significant issue in calculating the p-value if a t-test was used instead of the Wilcoxon Rank Sum Test. The p-value is 0.0014 which is significant less than 0.05 which means the distribution of homework averages for online and on-campus are significantly different.

Table 3: Summary of Homework Average

Homework Average Online	Homework Average on-Campus
Sample Size: 110	Sample Size: 284
Minimum: 42.64	Minimum: 54.10
1 st Quartile: 90.54	1 st Quartile: 93.91
Median: 95.19	Median: 97.42
Mean: 91.93	Mean: 95.12
3 rd Quartile: 98.50	3 rd Quartile: 99.54
Maximum: 100.00	Maximum: 100.00

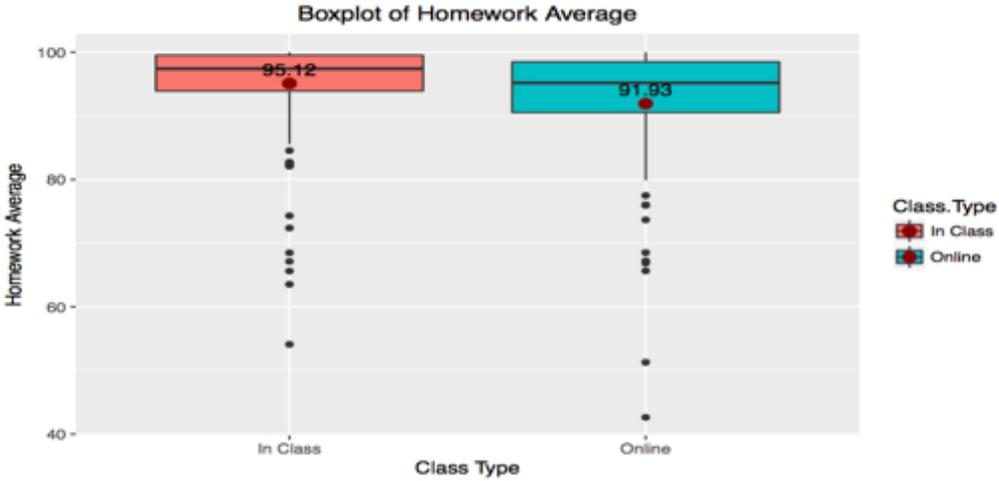


Figure 4: Boxplot of Homework Average

The next exploration was for number of attempts. For number of attempts, looking at Table 4 the number of attempts for quiz and highest score attempts for both environments have around two attempts or Figure 5 for only the overall quiz. However, for the p-value calculated for number attempts for a quiz gave a 0.04245 which means the distributions of number of attempts for quizzes are significantly different for online and on-campus course. However, if more data in the future is added to this research than the p-

value could possibly change where the two distributions might not be significantly different.

Table 4: Summary of Number of Attempt

Online Quiz Number of Attempts	On-Campus Quiz Number of Attempts	Highest Score Online Quiz Attempts	Highest Score on- Campus Quiz Attempts
Sample Size: 110	Sample Size: 284	Sample Size: 110	Sample Size: 284
Minimum: 1.000	Minimum: 1.000	Minimum: 1.000	Minimum: 1.000
1 st Quartile: 2.000	1 st Quartile: 1.000	1 st Quartile: 2.000	1 st Quartile: 1.000
Median: 2.000	Median: 2.000	Median: 2.000	Median: 2.000
Mean: 2.243	Mean: 2.118	Mean: 2.175	Mean: 2.055
3 rd Quartile: 3.000	3 rd Quartile: 3.000	3 rd Quartile: 3.000	3 rd Quartile: 3.000
Maximum: 5.000	Maximum: 8.000	Maximum: 5.000	Maximum: 8.000

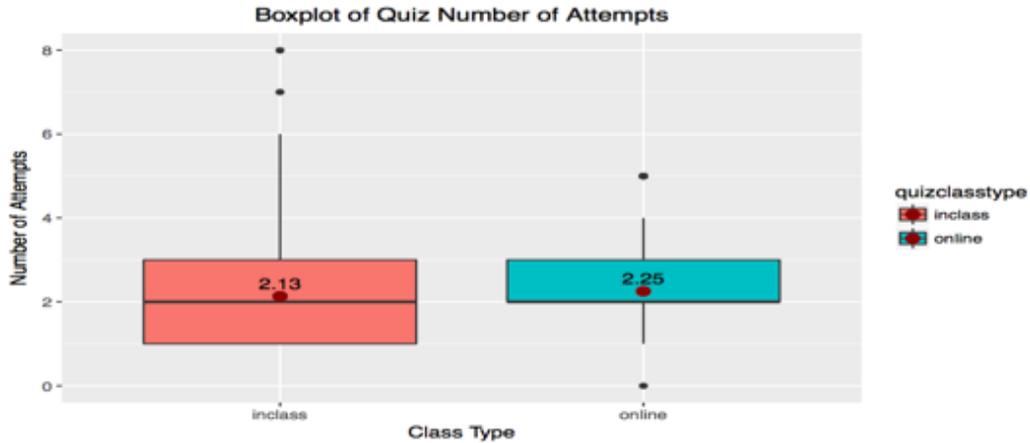


Figure 5: Boxplot of Number of Attempts

DISCUSSION

The results of comparison of on-campus and online courses in the introductory probability and statistics were expected in some cases while others were not. By looking at the p-values calculated when comparing online to on-campus data, the data shows the final

quiz average (p -value = 0.7615) and first attempt quiz average (p -value = 0.8280) don't have significantly different distributions. The assumption of why this occurs is due to the steps taken to learning the material by the homework and the first attempt quiz being similar to the other quizzes the student takes. However, for the homework average, on-campus and online have significantly different distributions (p -value = 0.0014). A student in an on-campus course has several resources to receive help like the learning center which assist with understanding the materials or attending the professor's office hours for more understanding. Also, in some courses taking on-campus may benefit more due to the learning center providing a supplemental instructor to hold sessions outside of the class to cover the material discussed in class and answer questions a student may have over the homework. These beneficial resources may be the reason for the homework average to be higher for on-campus. When comparing the student's performance for on-campus and online, looking at the number of attempts provides evidence of if an online student needs more attempts due to the lack of resources or if both on-campus and online is comparable. The p -value is 0.04245 which means the number of attempts for on-campus and online have significantly different distributions. However, since the value is close to the 0.05 the accumulation of more data could change the conclusion.

The two major data limitations which prohibited further analysis is the days before due date between online and on-campus due to the length of the semesters varying greatly. The other limitation is the data sizes for the quiz comparisons being small in size which could be improved by adding more data. Other than these two data limitations, the comparisons of online and on-campus are comparable; however, if a student should choose to take the class online or on-campus, on campus would be the better choice since the homework average is higher and quiz comparisons are slightly higher for all, but one quiz and the student has several resources to receive help from.

Collecting more data would be the first step in the further research. Since all the data in this research was from one statistics professor it would be interesting to see if the same conclusion still holds after getting data from other teachers and the addition of more semesters for online and on-campus. In addition, including more variables like how long a student took on taking the assignment, the attendance of the student, number of times a student logs on into MyStatLab, and whether a student took the assignment in one session

or multiple sessions. Adding the variable of how long a student took taking the assignment would help judge if a student actually truly attempted the first attempt or if the student looked at the first attempt as a view of what the quiz will cover. Analyzing this data will help determine if the first attempt average is accurate. For on-campus, the attendance of a student would help analyze if a student who attends class receives a higher quiz and homework grade over students who less frequently attend class. Analyzing the number of times a student login to MyStatLab would help determine if a student only logs on the days the assignment are dues or the day the student first attempt the assignment. Adding these variables and more data would better enhance the continuing research to get a more accurate comparison of the environments based on the student's performance.

REFERENCES

Hastie, Trevor, Tibshirani, and Friedman. 2001. *The Elements of Statistical Learning*. New York: Springer.

Hogg, Tanis, and Zimmerman. *Probability and Statistical Inference*, Ninth Edition. Pearson.

Wackerly, Mendenhall, and Scheaffer. *Mathematical Statistics with Applications*, Seventh Edition. Brooks/Cole Cengage Learning.

Wickham, H. *ggplot2. Elegant Graphics for Data Analysis*.