# Quizzes as a measure of instructors' style of teaching

Elena Shevchenko

## Introduction

Quizzes is an essential part of students' academic assessment, which can be used to measure growth in knowledge and skills. However, education is interaction between students and instructor and, therefore, success or failure of a student is related to instructor's method of teaching. Some instructors choose conservative approaches while others prefer to use loyal methods of teaching. These method may impact students' ability to understand information and develop skills, where quizzes show quantitative assessment of these processes. Thus, quizzes also can be considered as a measure of instructors' style of teaching. In this paper instructors' style of teaching will be explored based on the results of quizzes.

## Data

The data includes 293 average quiz scores for 5 weekly quizzes in Statistics and Probability course and exam results for the same course, which was conducted after the last quiz. Each student was assigned to one instructor (A, B, C or D). Instructor A was assigned to 69 students, instructor B - 64, instructor C - 90 and instructor D - 70. Additionally, students were divided into groups (1 through 13). Most of the groups contain 23 students, while the others contain 18, 22 or 24 students. Data set also includes extra information that can be used in further analysis and will not be discussed in this paper. Quizzes were graded by each instructors, while exams were graded by other experts where exams were assigned randomly to different graders.

## EDA

Descriptive statistics for each variable is presented below. Quiz grades scale is 0-10 with almost equal mean and median.

| | group | quiz_average | Exam_1 | Total | Grade |
|---|---|---|---|---|---|
| count | 281.000000 | 281.000000 | 281.000000 | 281.000000 | 281.000000 |
| mean | 7.209964 | 5.988968 | 74.227758 | 104.172598 | 71.892981 |
| std | 3.667722 | 2.359612 | 22.395585 | 30.204061 | 20.999297 |
| min | 1.000000 | 0.400000 | 10.000000 | 32.000000 | 17.495500 |
| 25% | 4.000000 | 4.200000 | 59.000000 | 83.000000 | 57.173500 |
| 50% | 7.000000 | 5.800000 | 78.000000 | 110.000000 | 76.341000 |
| 75% | 10.000000 | 7.800000 | 93.000000 | 128.000000 | 89.332000 |
| max | 13.000000 | 10.000000 | 105.000000 | 155.000000 | 104.167500 |

Figure 1: Descriptive statistics

It can be seen from the table below that instructor A has the lowest quiz average , while instructor C has the highest quiz average. Immediate conclusion is that teaching methods of instructor A do not work on the students, while instructor C has more efficient approaches in teaching their students. However, there may be some other explanations of such results. One is that instructor C is more loyal in grading than instructor A. Another explanation is that it might happened that instructor A had students with lower level of mathematical skills who did not have experience in statistics before, while instructor C had more experienced students in this area. To control for difference between students' abilities to learn the material SAT scores can be considered. However, no extra data is available to instructors of this course. Moreover, the number of students assigned to each instructor and randomization of assignment allow to conclude that students should be approximately equally distributed between instructors according to their maths abilities.

| instructor | quiz_average |
|---|---|
| A | 3.993846 |
| B | 5.216949 |
| C | 8.086667 |
| D | 5.786567 |

Figure 2: Quiz average by instructor

Figure 3 shows that even though quizzes average was the lowest for instructor A, exams average is not the lowest among instructors. Additionally, instructor C has the highest exam average. The difference in exam grades for instructors A, C and D is negligible, while instructor B has significantly low exam average.

| instructor | Exam_1 |
|---|---|
| A | 75.276923 |
| B | 65.966102 |
| C | 77.300000 |
| D | 76.358209 |

Figure 3: Exam average by instructor

Boxplots presented below show that exam grades for three instructors A, C and D look similar. Boxplot for instructor B is different from others and the box is located slightly lower than boxes for other instructors. Half of students for all three instructors received grades between 55 and 95, while for instructor B half of students received grades between 50 and 80. It is clear that the minimum exam grade is the highest for instructor A. The range for each instructors' grade is presented below two boxplots and shows that instructor A has the lowest range, while instructor C has the highest range. Standard deviation is also lowest for instructor A, while instructor D has the highest standard deviation.

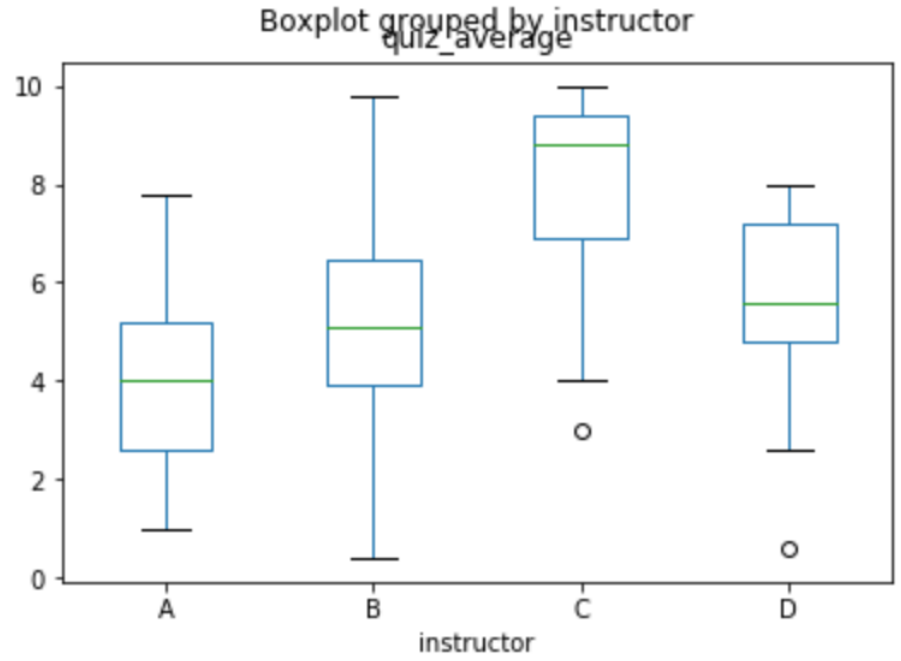


Figure 4: Quiz averages, boxplot

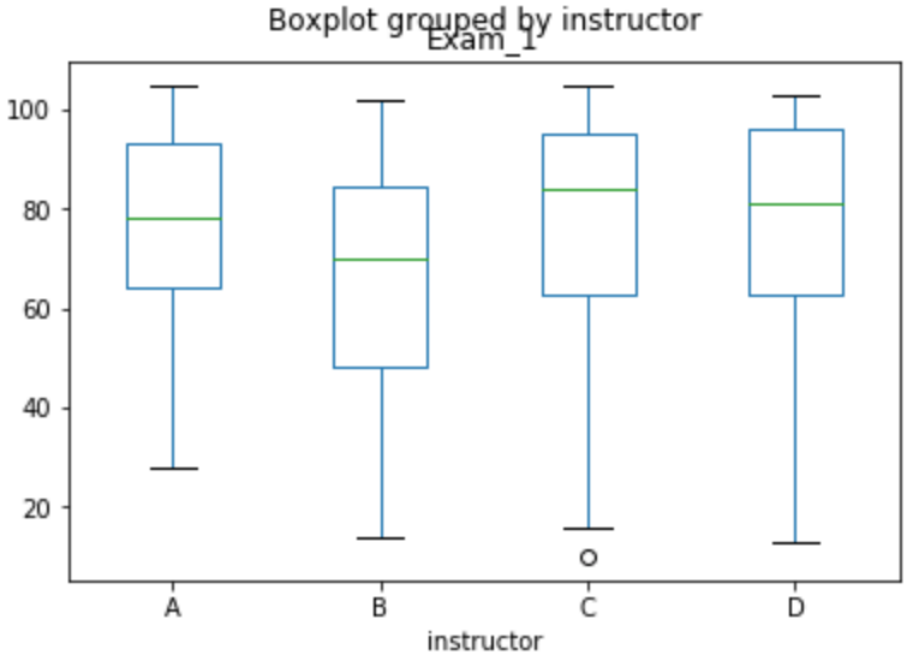


Figure 5: Exam grades, boxplot

| | Exam_1 |
|---|---|
| instructor | |
| A | 77 |
| B | 88 |
| C | 95 |
| D | 90 |

Figure 6: Range

| | | |
|---|---|---|
| | instructor | |
| Exam_1 | A | 21.529419 |
| | B | 22.176382 |
| | C | 22.256144 |
| | D | 22.330643 |

Figure 7: Standard deviation

## Analysis

ANOVA is an appropriate analysis here since the interest is to test whether there is difference between quiz averages among instructors controlling for exams that students received. To use ANOVA it is necessary that the relationship between independent and dependent variables is linear. Both Figure 6 and correlation matrix show that there is linear positive relationship between two variables. Correlation matrix by instructor shows that the highest linear relationship between quiz grades and exam grades is for instructor A, while the lowest correlation coefficient is for instructor B.

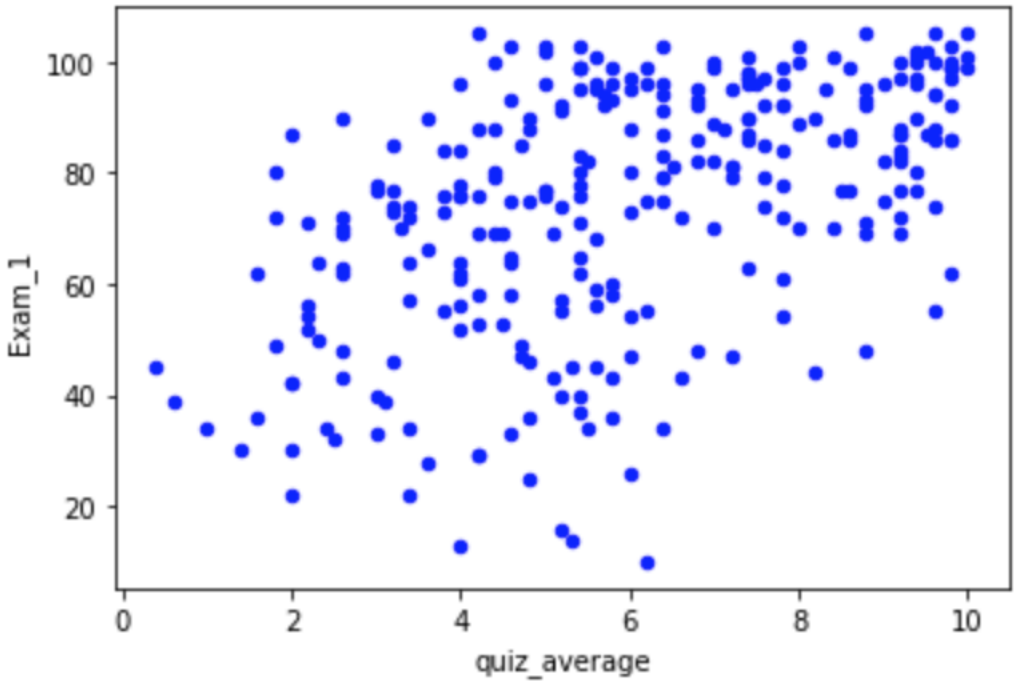


Figure 8: Scatterplot

| | quiz_average | Exam_1 |
|---|---|---|
| quiz_average | 1.000000 | 0.513822 |
| Exam_1 | 0.513822 | 1.000000 |

Figure 9: Correlation matrix

| instructor | | quiz_average | Exam_1 |
|---|---|---|---|
| A | quiz_average | 1.000000 | 0.711916 |
| | Exam_1 | 0.711916 | 1.000000 |
| B | quiz_average | 1.000000 | 0.568907 |
| | Exam_1 | 0.568907 | 1.000000 |
| C | quiz_average | 1.000000 | 0.658733 |
| | Exam_1 | 0.658733 | 1.000000 |
| D | quiz_average | 1.000000 | 0.609444 |
| | Exam_1 | 0.609444 | 1.000000 |

Figure 10: Correlation matrix by instructor

Figure 11 presents the results of ANOVA for quiz averages. The conclusion of ANOVA is that means of all quiz averages are not all equal since null hypothesis is rejected due ti small p-value. Post hoc test TukeyHSD is conducted to determine pairs that are not equal. According to Figure 12, all pairs of means cannot be considered equal except for instructor B and D.

| | sum_sq | df | F | PR(>F) |
|---|---|---|---|---|
| instructor | 692.673301 | 3.0 | 73.82737 | 4.065494e-35 |
| Residual | 866.302500 | 277.0 | NaN | NaN |

Figure 11: ANOVA for quiz average

Multiple Comparison of Means - Tukey HSD, FWER=0.05

| group1 | group2 | meandiff | p-adj | lower | upper | reject |
|---|---|---|---|---|---|---|
| A | B | 1.2231 | 0.001 | 0.4011 | 2.0451 | True |
| A | C | 4.0928 | 0.001 | 3.3488 | 4.8369 | True |
| A | D | 1.7927 | 0.001 | 0.9969 | 2.5885 | True |
| B | C | 2.8697 | 0.001 | 2.104 | 3.6354 | True |
| B | D | 0.5696 | 0.2737 | -0.2465 | 1.3857 | False |
| C | D | -2.3001 | 0.001 | -3.0377 | -1.5625 | True |

Figure 12: TukeyHSD for quiz average

ANOVA results for exam grades are presented below. Null hypothesis is rejected due to small p-value and, therefore, some means can be considered different. TukeyHSD test shows that almost all means for exam grades can be considered equal except means between instructor B and instructor C, instructor B and instructor D. Albeit, there is no statistically significant evidence to conclude that means between instructor B and instructor A are different, it can be seen that the upper bound of confidence level for means between instructor B and instructor A is slightly greater than zero, meaning that their means are close to become different.

| | sum_sq | df | F | PR(>F) |
|---|---|---|---|---|
| instructor | 5252.172914 | 3.0 | 3.587304 | 0.014247 |
| Residual | 135185.250573 | 277.0 | NaN | NaN |

Figure 13: ANOVA for exam

Multiple Comparison of Means - Tukey HSD, FWER=0.05

| group1 | group2 | meandiff | p-adj | lower | upper | reject |
|---|---|---|---|---|---|---|
| A | B | -9.3108 | 0.0907 | -19.5787 | 0.957 | False |
| A | C | 2.0231 | 0.9 | -7.2717 | 11.3179 | False |
| A | D | 1.0813 | 0.9 | -8.86 | 11.0226 | False |
| B | C | 11.3339 | 0.0128 | 1.7686 | 20.8992 | True |
| B | D | 10.3921 | 0.0438 | 0.1974 | 20.5868 | True |
| C | D | -0.9418 | 0.9 | -10.1557 | 8.2721 | False |

Figure 14: TukeyHSD for exam

## Conclusion

Analysis of quiz averages indicates that all means are different except for one pair (instructor B and instructor D) with sample averages 5.22 and 5.79, respectively. In spite of the fact that instructor A has the lowest quiz average, exam average can be considered the same as for instructors C and D, who have the highest quiz averages. It is likely that instructor A is more conservative and demanding with grading than other instructors, which can impact both students' progress. Additionally, instructor A has the highest correlation coefficient between quiz grades and exams (0.71), which can be considered as the most fair and unprejudiced quiz grading. In this case students may loose points for typos or unclear solution, which stimulates them to learn more and to improve their skills to receive better grades. Instructor B, on the contrary, has higher quiz grades along with the lowest exam average, which is significantly differs from other instructors. One possible explanation is that instructor B is more loyal to students in grading. Instructor B may not mark some minor errors or incomplete solutions, which may lead to students' unawareness of their mistakes. This can affect learning process and cause lower grades on the exam.