Comparison of Student Test Performance Over a Term Based on Pre-Test Score and Teaching Format

McCord Murray

February 23, 2022

Abstract

The goal of this experiment is to evaluate the improvement of students' test scores over the duration of a term based on how they scored on the pre-test and their learning environment. A bootstrapping method will be used to randomly test the frequency in which the difference of the students' gains is greater than the created groups. After using bootstrapping to generate a p-value it was found that students who scored lower on the pre-test and students in the pilot-group demonstrated statistical differences from the other distributions too unlikely to accept that being in these groups did not positively impact the gain.

Background and Significance

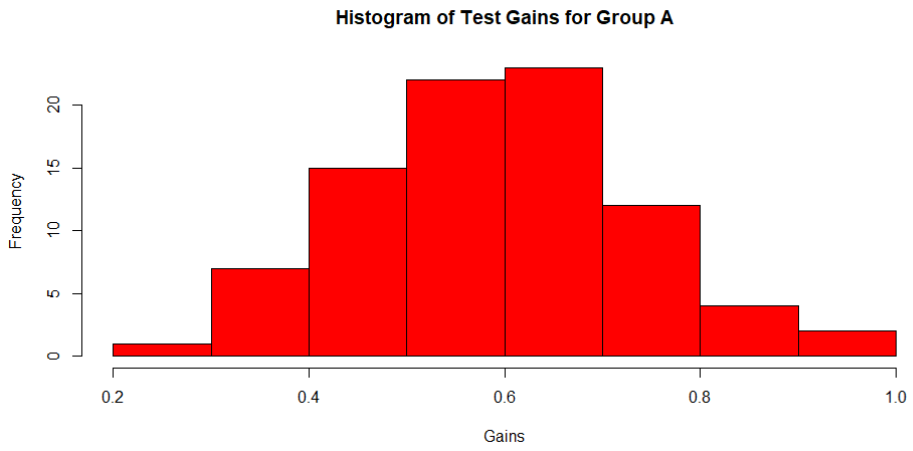
In education, a student's overall performance is used to evaluate their grade at the end of a term. The most common way of doing so is through standardized tests. One of the major issues with the standardized testing method is that the final grade does not reflect how much the student has improved over the duration of the term, meaning that many students are awarded a grade that does not accurately represent their effort. Additionally, the method in which the students are taught can also have an affect on their performance. In order to better evaluate how much a student has improved over a term, the students were given a 'pre-test' at the beginning of every term, and a similar 'post-test' at the end. In addition to this, another group of students were split into a group given a more traditional lecture style with a traditional textbook, and a group more focused on student engagement with a modified text to measure how teaching methods affect student improvement. The goal of this experiment is to evaluate the improvement of students' test scores over the duration of a term based on how they scored on the pre-test and their learning environment. The hypothesis is that students who scored lower on the pre-test improved more than students who scored higher on the pre-test, and students who learned in the modified learning environment improved better than students who learned in the traditional format.

Methods

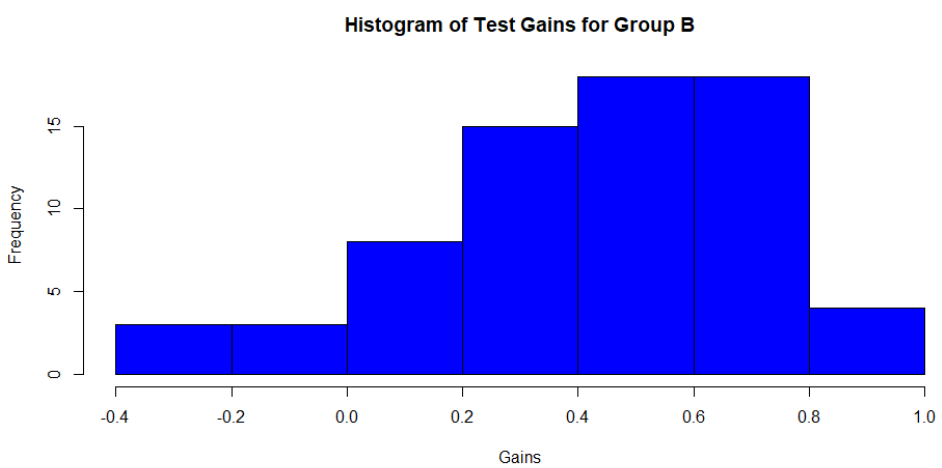
In order to test the hypothesis, a variable called 'gain' needs to be created to measure improvement. The variable gain will be defined as (post-test score - pre-test score)/(1 - pre-test score). To measure whether or not students who scored lower on the pre-test have a better gain on average than students who scored higher on the pre-test, students were separated into students who scored below or equal to the median score (Group A), and students who scored higher than the median score (Group B). A bootstrapping method will be used to randomly test the frequency in which the difference between a pseudo-Group A and a pseudo-Group B has a higher difference in mean of their gains than the actual Group A and Group B. The output of the bootstrapping method will be a p-value, which tells us the probability of the difference in the mean of the gains occurred assuming being Group A has the same effect on gain as being in Group B. This same bootstrapping method can be applied with the traditionally taught students (Group T) and the students in the modified curriculum, also called the pilot group (Group P).

Results

The histogram and descriptive statistics for the distribution of the gains of Group A, students who scored below or equal to the median on the pre-test is the following, and Group B, students who scored above the median are the following:

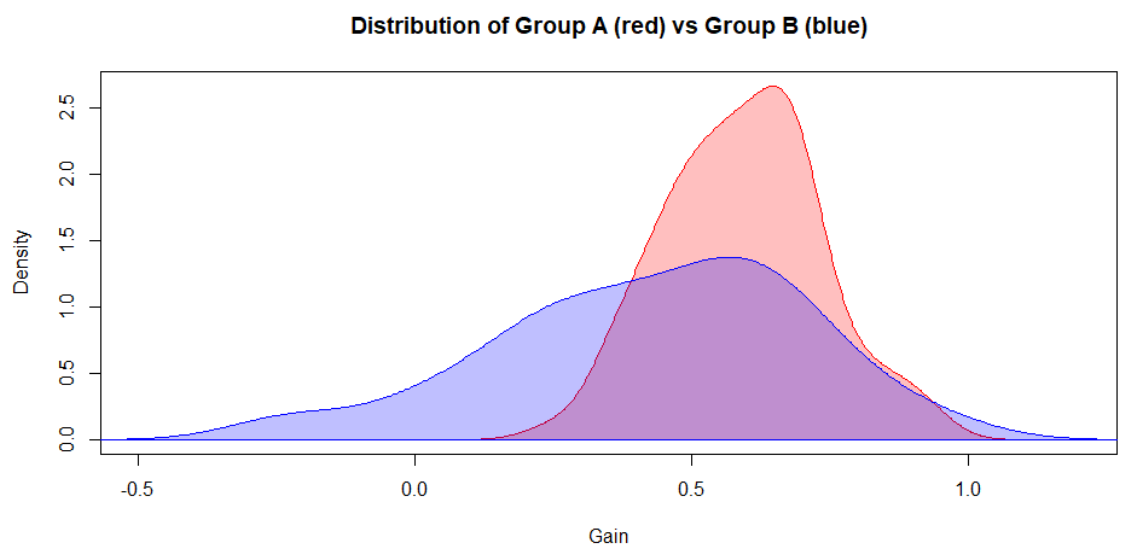


Descriptive Statistics	Gain of Group A
Mean	0.5914204
Median	0.5862069
Standard Deviation	0.1404466
Skewness	0.07600833
Kurtosis	2.774441
5-Number Summary	0.2456, 0.4951, 0.5862, 0.5914, 0.6845, 0.9286



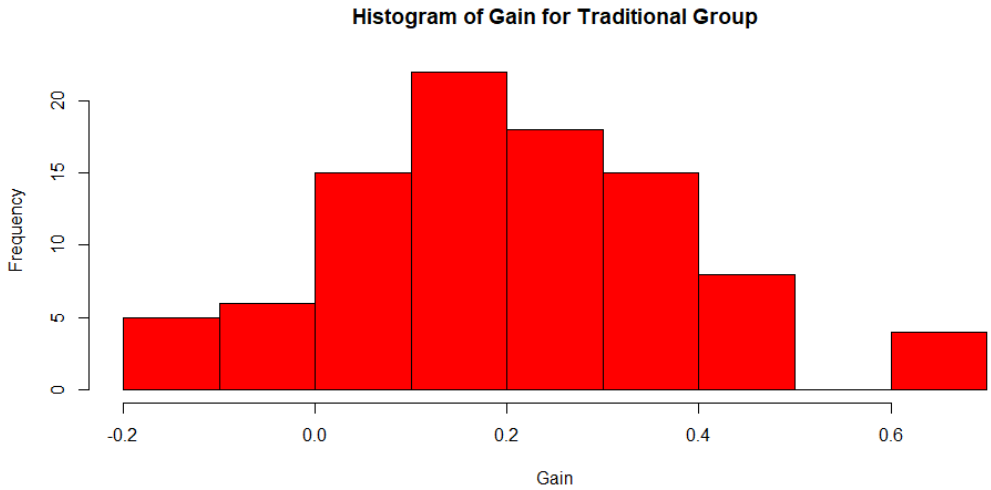
Descriptive Statistics	Gain of Group A
Mean	0.4285413
Median	0.4444444
Standard Deviation	0.2768818
Skewness	-0.4397681
Kurtosis	2.804831
5-Number Summary	-0.25, 0.2453, 0.4444, 0.6250, 0.9730

Overlaying the two distributions on top of each other (Group A: Red, Group B: Blue), it is observed that the peak of the gains in Group A is much higher than Group B.

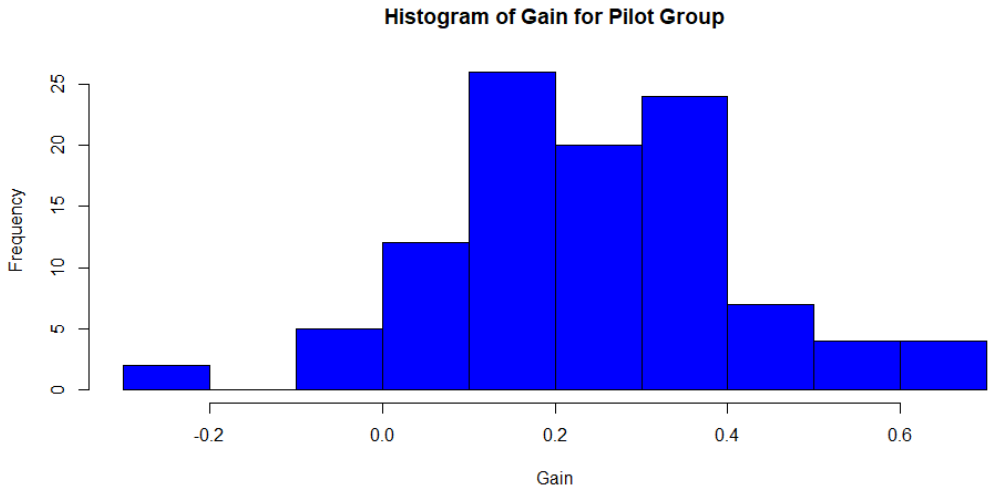


The difference in the means of the gains of Group A and Group B is 0.162879. Using bootstrapping on the gains of Group A and Group B to test how frequently the difference in means of the pseudo-Group A and pseudo-Group B were greater than 0.162879 yielded a p-value of 0.000001 or 1/1000000.

The same process was done for the Traditional Group and the Pilot Group. The histogram and descriptive statistics for the distribution of the gains of Group T (Traditional), students who were taught with a traditionally styled lecture and textbook, and Group P (Pilot), students who were taught with a greater focus on student engagement and used a modified text.

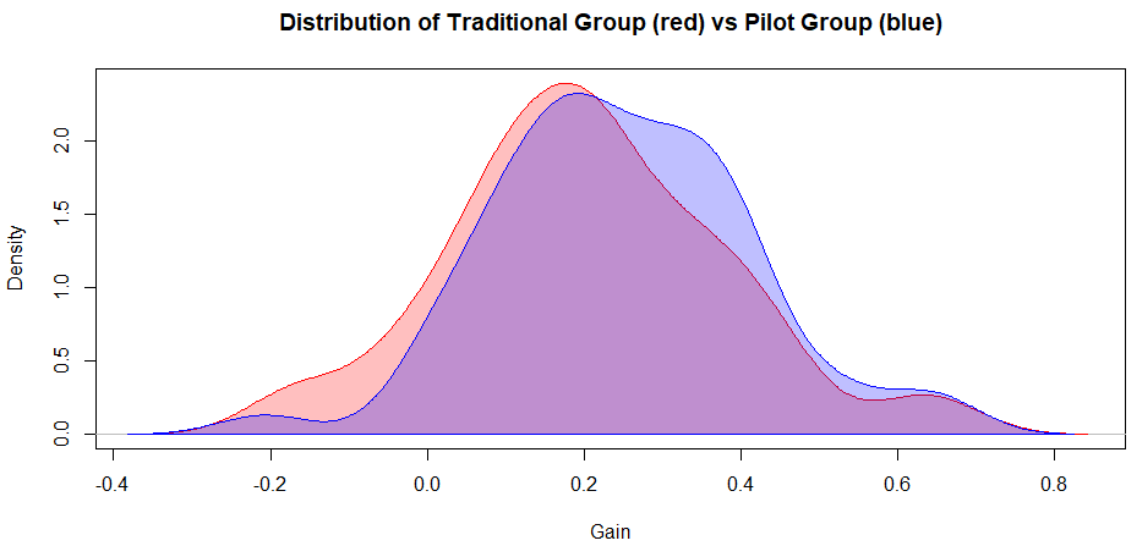


Descriptive Statistics	Gain of Group T
Mean	0.2042765
Median	0.1944444
Standard Deviation	0.177572
Skewness	0.2355234
Kurtosis	3.224763
5-Number Summary	-0.19444, 0.08861, 0.19444, 0.30864, 0.66324



Descriptive Statistics	Gain of Group P
Mean	0.2465833
Median	0.2437146
Standard Deviation	0.1641452
Skewness	0.1215444
Kurtosis	3.439531
5-Number Summary	-0.2069, 0.1429, 0.2437, 0.3704, 0.6486

Overlaying the two distributions on top of each other (Group T: Red, Group P: Blue), it is observed that the pilot group has more of a less positive skewness than the traditional group.



The difference in the means of the gains of Group P and Group T is 0.04230673. Using bootstrapping on the gains of Group P and Group T to test how frequently the difference in means of the pseudo-Group A and pseudo-Group B were greater than 0.04230673 yielded a p-value of 0.041996 or 41996/1000000.

Conclusion

The goal of this experiment was to evaluate the improvement of students' test scores over the duration of a term based on how they scored on the pre-test and their learning environment. The hypothesis was that students who performed worse on the pre-test would have a higher gain than students who performed better, and that students learning in a more interactive environment (pilot group) would have a higher gain on average than students learning in a more traditional environment. After using bootstrapping to generate a p-value, assuming each scenario had no impact on the gain, it was found that students who scored lower on the pre-test and students in the pilot-group demonstrated statistical differences from the other distributions too unlikely to accept that being in these groups did not positively impact the gain. It is not surprising that being in Group A results in a higher gain since those students have more room to improve than students who place in Group B. As for the pilot group having a higher gain than the traditional group, this result posits that a more interactive program increases student performance over the term, though no details were provided on how the pilot group was more interactive or how the text was modified from the traditional text.