Gain Scores for Two Groups

Nick Collins

UMass Dartmouth

March 4, 2022

Abstract-

Pre-test and post-test scores can give great insight to teachers. After using some sort of study guide/technique the changes in these scores can change dramatically and show just how well the study guide/technique worked. Using descriptive statistics, graphs, differences in means and t scores we can get some insight into how the groups are different. There were two sets of data for this project, each of which had their own two groups. Using these two groups in each data set we can try to show there is or isn't enough evidence to support the claim that the two sets are statistically similar.

Introduction-

Using two different data sets we want to examine if they are statistically similar, or statistically significantly different. The first data set consists of pre-test and post-test percentage grades (scaled to lie between 0 and 1) for a cohort of 155 pre-service teachers taking a college mathematics course. The second data set consists of mathematics pre-test and post-test scores for a cohort of students in a mid-west university. The cohort of students is split into two groups, traditional and pilot. When we apply some simple descriptive statistics, we can see the differences between the groups and the data sets. This can be done through looking at means, difference of means and p values. We can also clearly see the distribution for the difference in means when plotting them in a histogram.

Methods-

There were two data sets used for this project. The first data set was 155 rows and two columns. The second data set was split into two groups, traditional and pilot. The traditional data was 93 rows and 2 columns. The pilot data was 104 rows and 2 columns. The two columns in each of the data sets were the pre-test and post-test percentage grades.

The first data set needed to be split into two groups based on the pre-test median. In order to do this I had to first find the median and then sort the data. After this was done, I could see where the pre-test scores were less than the median and assigned them to group A. Group B were the remaining values, which were greater than or equal to the pre-test median. After splitting the data into the two groups, scores needed to be computed. Looping through the data frames and assigning the gain scores to the correct rows. The formula for the gain scores was given on the assignment and was easy to calculate using a for loop. The next step was to randomly assign scores to pseudo group A/B. These assignments were to be done at random and then the difference in means for group A and group B was to be calculated and added to a list to be plotted. The assignments had to be done many times (10,000 to 1,000,000 different times) so for this I used a while loop and a counter variable. After this loop was completed, I had it append the differences to a list that could be plotted to see the distribution of the differences in means when randomly assigning groups A and B.

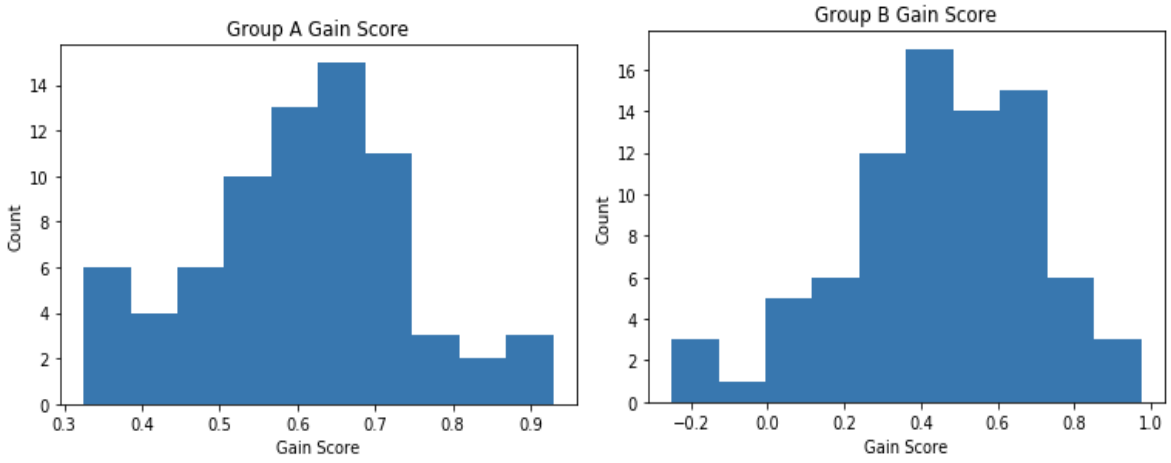
The second data set was similar to the first data set, except the second data set was already split up into the two groups that needed to be analyzed. The first group was the pilot group who “were

taught with a modified text that emphasized greater student engagement." The second group in this data set was the traditional group and they "were taught by a lecture style method using existing texts." After getting these two groups separated into different data frames, the next step was calculating the gain scores, like they were calculated in the first data set.

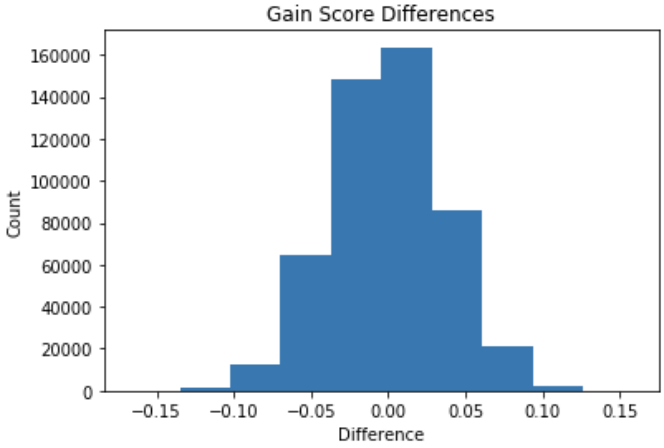
One of the last tasks that can give us some insight into the data is to perform a t-test on the two sets. To do this I needed to find the standard deviation and variance of the groups. For the first data set we needed to take the variance and standard deviation for group A and B. The same was true for the second data set, where we needed to get the variance and standard deviation for the pilot and traditional groups. Once we find these, we can get the t score which will give us a p value for each set.

Results-

For the first set of data the first task was to apply some simple descriptive statistics. The mean gain score for group A was 0.603735, the mean pre-test score was 0.282740 and the mean post-test score was 0.716438. Group B's gain score was 0.443400, the mean pre-test score was 0.573659 and the mean post-test score was 0.771463. So right away we see group A has a higher gain score average and we would expect to see this in the histograms.



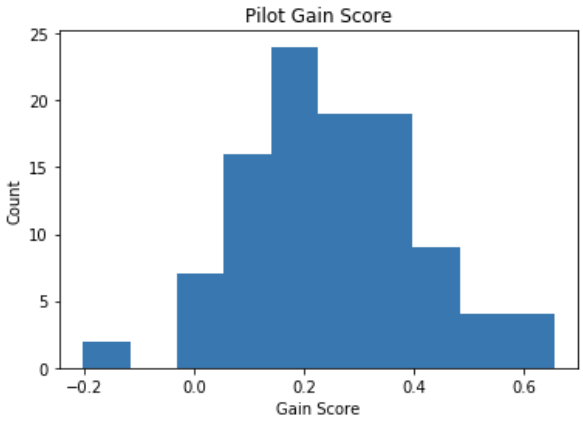
Looking at the two histograms we do see group A's gain scores peak around 0.6 to 0.7 and group B's gain scores peak around 0.4 to 0.6. Now we can assign pseudo groups A and B through the use of a loop and calculate the difference in means for the two groups. I decided to repeat this



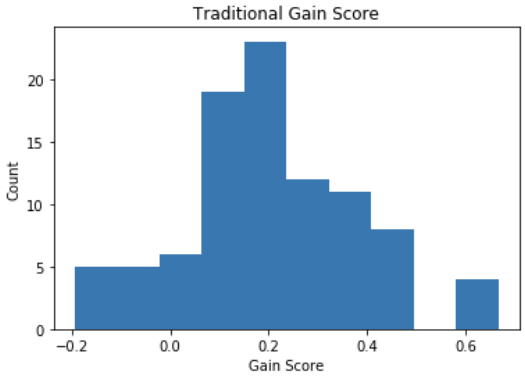
500,000 times and this is the output I got.

Here we see a decrease in the difference in means, which is good to see. When assigning the groups at random we should expect the difference in means to be smaller than when groups were assigned by the median pre-test score. After calculating the standard deviation and variance for group A and B we can compute a t-test and get a p value. Doing this t-test gave me the t score = 4.84723.

For second data set, the mean gain score for the pilot group was 0.246389, the mean pre-pilot score was 0.207019 and the mean post-pilot score was 0.401163. The mean traditional group gain score was 0.204823, the mean pre-trad score was 0.182296 and the mean post-trad score was 0.349087. When plotting the distributions of the gain scores for the pilot and traditional



groups we get the following.



The two groups have a similar distribution, but we see the traditional group has a mean gain score that is a little smaller than the pilot group. After calculating the standard deviation and variance for the pilot group and the traditional group we can compute a t-test and get a p value. Doing this t-test gave me the t score = 1.69817.

Conclusion-

After performing the analysis on the two data sets, we see some interesting results. Talking about the t scores first we see the first data set had a t score of 4.84723 and the second data set had a t score of 1.69817. The first data set has a much higher t score than the second data set, although the second data set also has a relatively high t score. The larger the t score the more variation and differences exist between the two groups of data. These two t scores show the similarities and differences between the same data sets. In other words, the t scores aren't comparing the two data sets to each other, rather they are comparing group A and B in the first set and pilot and traditional in the second set. With two t scores so high we can't prove they are statistically different, but these two scores give us some insight that it would be highly unlikely the two are statistically similar.

Looking at the histograms for gain scores amongst the two groups we can eyeball some of the statistics and get a rough estimation on what the data is doing. Looking at descriptive statistics we can see the true values for each group. The first data set had group A and B and group A had a higher average gain score than group B. This makes sense logically speaking because group A was all the pre-test scores below the median pre-test score. Since group A had lower pre-test scores this means they had more room to improve on the post-test. They had higher gain scores, but they didn't have a higher post-test average. This also makes sense because we should see a higher average post-test score for those who did better on the pre-test. After looping 500,000 times and assigning pseudo group A and B we got to see the difference in means for each

iteration and plotted them in a histogram. This histogram shows a decreasing trend in the difference of means. We see a somewhat normal distribution where it peaks around zero and then flattens out evenly as we get towards –0.1 and 0.1.

The second data set groups are more similar to one another than the first data set groups (hence the lower t score for data set two). We see the two distribution histograms for the gain scores and just looking at them, they are somewhat similar in distributions. With descriptive statistics we see the pilot group has a mean gain score of 0.246389, while the traditional group had a mean gain score of 0.204823. The other difference we see is in the pre and post scores for the pilot and traditional groups. The pilot group had a pre-test score of 0.207019 while the traditional group had a pre-test score of 0.182296. These two are very similar and should be expected to be similar. The pilot group had a post-test score of 0.401163 while the traditional group had a post-test score of 0.349087.

References-

MTH 332 Mathematical Statistics Spring 2022

https://mth332.files.wordpress.com/2022/01/mth_332_spring_2022.pdf