

Teaching Methods Analysis

Carly Lundgreen

September 28, 2019

Introduction

In educational settings, it is important to consider the effectiveness of various teaching methods and adjust these methods when appropriate. Educators must consider their students' responses to the material being taught and the delivery of this material. This analysis will focus on a comparison of the effects of two different teaching methods, A and B, on subjects' test scores. Our goal is to determine whether these two methods differ in their teaching effectiveness, where 'teaching effectiveness' is measured by average test scores.

Statistical Hypotheses

In order to determine whether the effects of teaching methods A and B differ from each other, we will establish the following hypotheses:

$$H_0 : \mu_A = \mu_B$$

$$H_A : \mu_A \neq \mu_B$$

Where μ_A represents the mean test score for the group that received instruction using teaching method A, and μ_B represents the mean test score for the group that was instructed with method B.

Methodology and Procedures

The strategy for this analysis is to first perform a standard two sample t-test to compare the average test scores of each group. This test will allow us to reject or fail-to-reject the null hypothesis (H_0) stated above by comparing the calculated p-value to an alpha level that we will specify below. To augment this analysis, we will also perform a Mann-Whitney U test, the non-parametric equivalent of the two sample t-test. The hypotheses for this test are as follows:

H_0 : "The distribution of all test scores for groups A and B are equal

H_A : The distribution of all test scores for groups A and B are not equal

By comparing the p-value for this test to our significance level, we can then reject or fail-to-reject the non-parametric H_0 .

Level of Significance

We will set the level of significance for this analysis as $\alpha = 0.05$.

Data Description and Exploratory Data Analysis

These data were gathered from two groups of students. Each group was randomly assigned to a teaching method, either A or B, and received instruction using that teaching method for 3 weeks. At the end of the trial period, each subject in the experiment then took the same test under a time limit of 45 minutes. The test scores of each subject were recorded. The data. There were 35 subjects in each group. The data used in this analysis had 70 rows and 2 columns: teaching method and test score.

Exploratory Plots

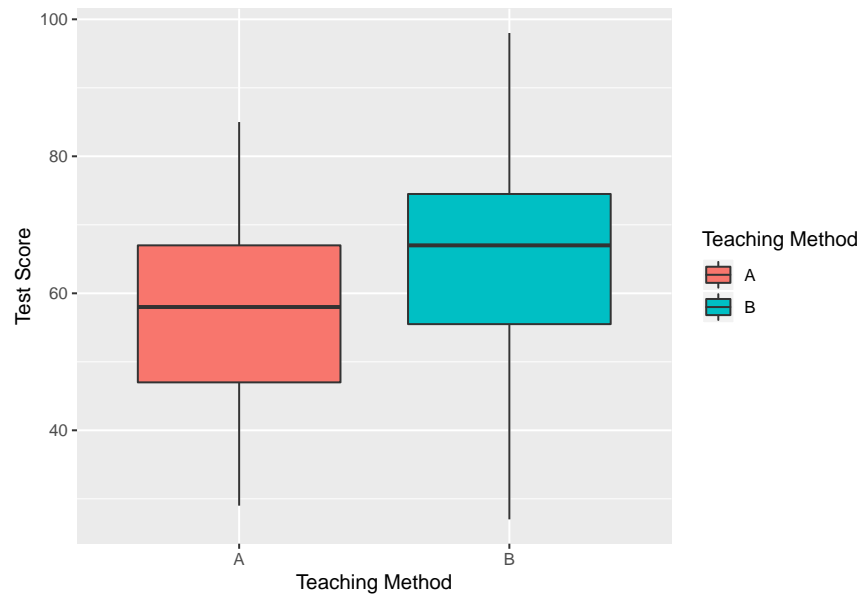


Figure 1: This figure depicts two boxplots that display the distributions of test scores for groups A and B.

These boxplots show that method B led to a wider range of test scores than did method A. We can also see that the median test score for method B is about 10 points higher than the median test score for method A.

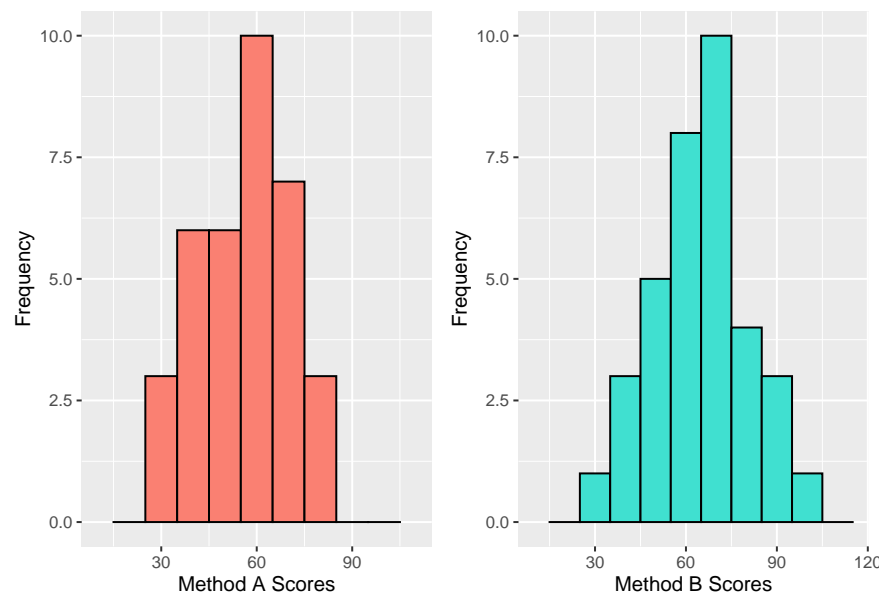


Figure 2: These plots display histograms of the test scores for teaching methods A and B.

Looking at these two histograms, both appear relatively symmetric and normal in shape as no significant skewness appears in either of the plots.

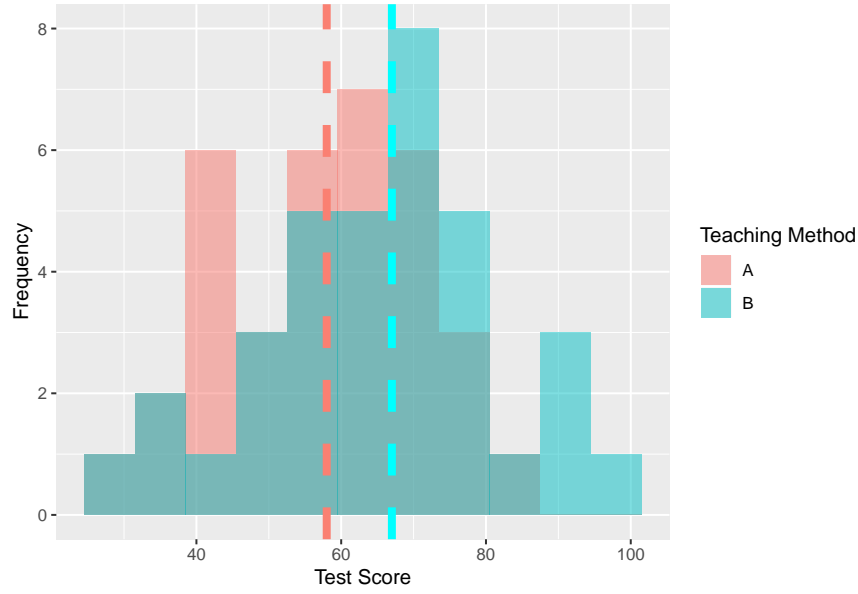


Figure 3: This histogram overlays the test scores from the subjects for groups A and B, and the median scores of both groups are also shown by dashed lines.

Summary Statistics

Teaching Method	Count	Mean	SD	Median	IQR	Min	Max
A	35	57.20	13.53	58	20	29	85
B	35	64.94	16.23	67	19	27	98

Group A had a mean test score of 57.2 while group B have a mean test score of 64.9428571. This is reasonable given that the maximum score for group B, which is 98 while group A have a maximum score of only 85. As mentioned previously, group B appears to have more variability in test score than group A, and this is confirmed when looking at the standard deviations. Score for group B have a standard deviation of 16.2262367 while group A scores have a standard deviation of 13.5338356.

We can also see that the median score for group A (58) is very close to its mean, which is the case in a theoretical normal distribution, so the scores for group A are likely distributed in a similar way. Group B has a few test scores that pulled its mean down below its median of 67, but the values are still relatively close together.

Finally, we will check the assumptions that are needed before performing a two sample t-test. A Mann-Whitney U test does not assume any distribution upfront, so we will proceed with that test either way. We confirmed that teaching methods were distributed randomly between the two groups of students to avoid biasing the results of the test. We also observed that the sample test scores do not have significant outliers or skewness and appear relatively normal in shape. Finally, we can see that the standard deviations are close enough in value (neither standard deviation is twice as larger as the other) to merit a two sample t-test.

Results

Below are the results of a two sample t-test and Mann-Whitney U test on the test scores of groups A and B. 95% confidence interval estimates for the difference in means (for the two sample t-test) and the difference in location parameter (median for the Mann-Whitney test) are provided.

Test	Estimate	Test_Statistic	P-value	Lower	Upper
2 sample t-test	-7.74	-2.17	0.0337815	-14.87	-0.61
Mann-Whitney	-8.00	432.00	0.0339058	-15.00	-1.00

In order to prove whether the difference in means for the test scores was truly different than 0 (the means of the groups do not equal each other), we performed a two sample t-test. This test calculates a test statistic, t , and the p-value is then calculated to be the probability of actually obtaining this test statistic if the means for the groups were equal. The Mann-Whitney test does not compare the means of the populations, but rather the distributions themselves. First, the data are pooled together, ordered, and ranked. If the distributions are truly equal, then we will expect to see a large test statistic, U , which would in turn lead to a significant p-value. We performed these tests using the appropriate R functions: `t.test` and `wilcox.test`.

Conclusions

Provide and Explain Conclusions

The p-value of the two-sample t test was 0.0338 while the p-value for the Mann-Whitney test was 0.0339. Both of these p-values are, firstly, very close together, and secondly, less than our significance level of 0.05, which leads us to reject both null hypotheses. We can conclude that the mean test score for group A is significantly different than the mean test score for group B. A 95% confidence interval for the true difference in mean test scores ($\mu_A - \mu_B$) was reported to be (-14.87, -0.61). This indicates the mean test score for all students who are instructed using teaching method B is higher than the mean test score for students instructed using method A, with 95% confidence.

The results of the Mann-Whitney test indicate that the distribution of test scores for students instructed with teaching method A differs from the distribution of test scores for students instructed with method B by a location shift of μ (this location shift is not a difference in averages, however. It is a comparison between the medians. A 95% confidence interval for the location parameter, μ , is calculated to be (-15,-1). This indicates that the test scores for group A are shifted to the left of the test scores for group B. When comparing the medians for groups A and B, 58 and 67 respectively, we can see that the median score for group A is less than the median score for group B, so this result is believable.

Questions

The first question I would ask the client is whether it would be feasible to increase the sample size of the groups. While 35 subjects per group is enough to get some preliminary results, it would be interesting to see the results if 100 or more students took received each treatment. I would also be interested to know whether the students took the test at the same time and location, as well as the locations from which these students were gathered. I would like to look for potential lurking variables that may be biasing the results and potentially confounding the effects of the teaching methods on the test scores with other potential variable(s).

Nonparametric vs. Parametric Methods

The biggest difference between non-parametric and parametric methods is that non-parametric methods require less conditions to be checked before conducting a test. With a two sample t-test, we had to have reasonable evidence that the scores from both groups came from a normal distribution. The Mann-Whitney test requires no such condition. However, this check is relatively easy to do, especially with relatively small sample sizes. The results that we got from this dual analysis turned out to be almost exactly the same, and led to, essentially, the same conclusions. In this case, a two sample t-test would be have been sufficient to answer the question of whether the effects of the teaching methods were different.

Recommendations

I would recommend, first, a larger sample size. This would lead to an increase in power. The probability of correctly rejecting the null hypothesis, or hypotheses, would increase. I would also recommend a longer trial period than three weeks for the experiment. If the students are just getting used to a new teacher, assuming the trial period begins at the beginning of a school year, test scores could be affected by the teacher and confound the results. If the study was conducted at the beginning of the year, then students may have performed differently had they taken the test in the middle or at the end of the year.

Future studies may include follow-up analyses on the same subjects. If the subjects were in middle school at the time of the study, perhaps the teaching methods would have different effects on the same students when they reach high school. The same concept may apply to college students. Perhaps a control group could be added as well, to compare the two treatments to a baseline treatment that has already been in practice. This would allow us to compare treatments to a baseline and see if either treatment is “better” than the methods already in place.