Exercise 8

In this exercise, you wanted to examine whether there are differences in students' scores on the first and second tests before any method is taught. Your problem is a hypothesis test for the equality of means between two populations. Since the measurements for both methods were taken from the same experimental units, you conclude that your samples are dependent. This is a "Before-After" experiment. To answer your initial question, you should form the differences in students' scores between the first and second tests, transforming your problem into a one-population mean test. Specifically, you are testing whether the mean of the difference in scores between the first and second tests is equal to zero.

To use the t-test for a single population, you should satisfy the following assumptions for your sample:

1. The sample should be random.
2. It should not contain extreme values (outliers) beyond 10% of the data.
3. The sample should come from a population that can be adequately described by a normal distribution.

Based on the histogram (Boxplot) you provided, it appears that there are no extreme outliers in the sample values of the differences in students' scores between the first and second tests (see Boxplot 1).

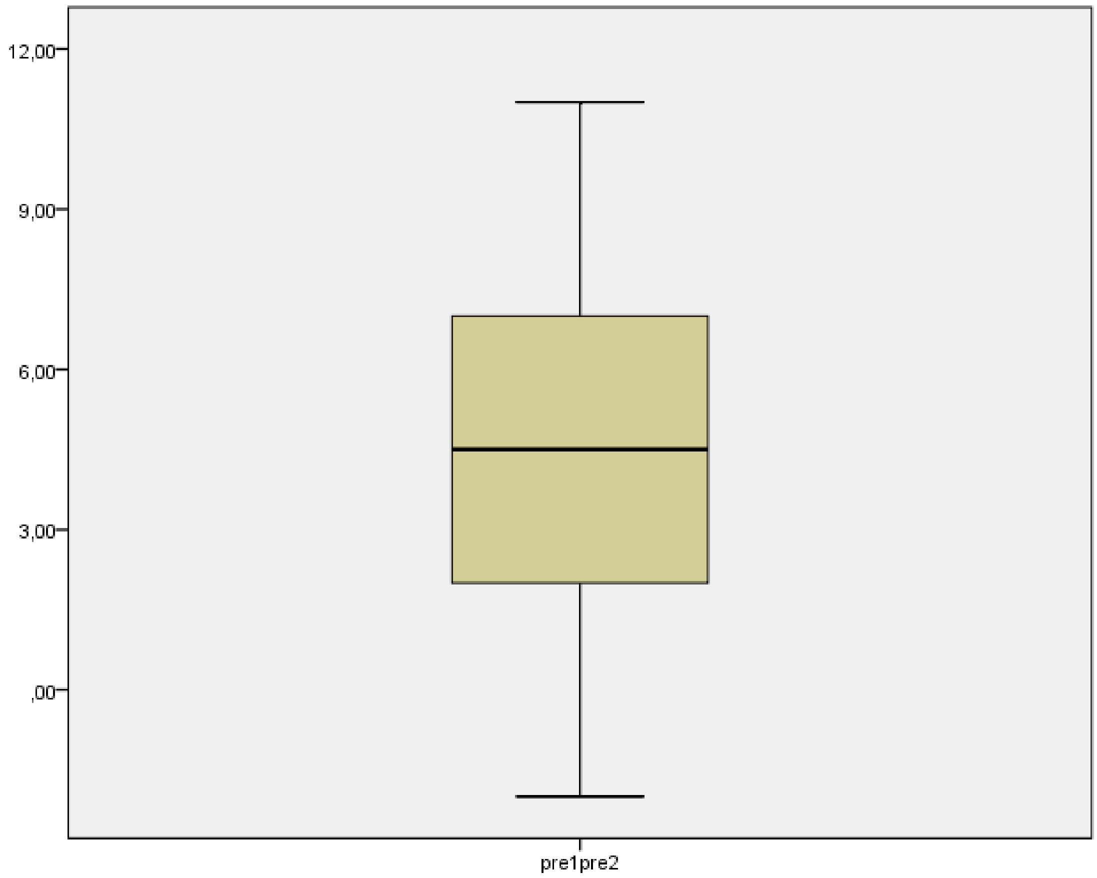
Next, you checked if the 66 sample observations of the difference in students' scores between the first and second tests come from a normally distributed population. The critical probability of the Shapiro-Wilk test is p-value=0.247, which means that you cannot reject the assumption of normality at a 5% significance level.

As a conclusion from implementing the t-test, it is evident that the scores of students on the first and second tests differ significantly ($p<0.001$). More specifically, the mean score on the first test is 4.68182 units better than on the second test (Table 1).

Table 1

		t			
		Mean	95% Confidence Interval of the Difference		Sig. (2-tailed)
			Lower	Upper	
Pair 1	Pre1 - Pre2	4,68182	3,92239	5,44125	,000

Boxplot 1



"We want to check if there is a statistically significant difference in the mean scores of two tests after teaching using three different methods. To use parametric statistical methods, the following assumptions need to be met:

1. Our samples should be randomly selected.
2. There should be no extreme outliers in the sample data for each population exceeding 10%.
3. Each population should be satisfactorily described by a normal distribution.

The first assumption, related to how we selected our samples, is satisfied.

The check for outliers in the sample of scores describing the teaching method 'Strat' showed that the observation with a serial number 52 is an extreme outlier (out of 22 available, representing less than 10%), and it is excluded from further analysis. The check for outliers in the sample of scores describing the teaching method 'DRTA' showed that the observation with a serial number 34 is an extreme outlier (out of 22 available, representing less than 10%), and it is also excluded from further analysis. The check for outliers in the sample of scores describing the teaching method 'Basal' showed that there are no extreme outliers (see histograms 1, 2, 3).

The assumption that the data representing the students' scores with the 'Basal' teaching method follows a normal distribution is not rejected (Shapiro-Wilk test p-value = 0.214).

The assumption that the data representing the students' scores with the 'DRTA' teaching method follows a normal distribution is not rejected (Shapiro-Wilk test p-value = 0.773).

The assumption that the data representing the students' scores with the 'Strat' teaching method follows a normal distribution is not rejected (Shapiro-Wilk test p-value = 0.616).

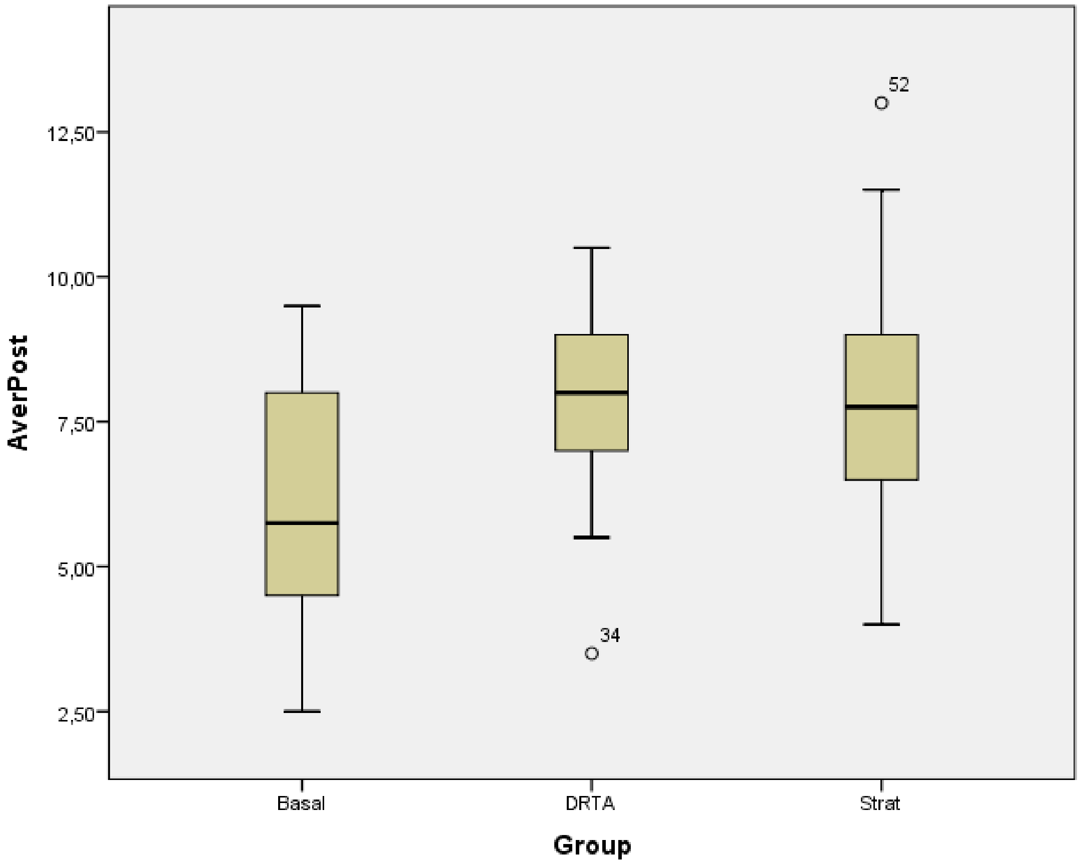
The hypothesis of equality of population variances for students' scores across the three teaching methods is rejected at a significance level of 5% (Levene's test p-value = 0.049).

Therefore, to test if there is a statistically significant difference in the mean population scores in the two tests after teaching, we will use the Welch's test with a significance level of 5%.

There is a statistically significant difference in the mean population scores in the tests after teaching (Welch's test p-value = 0.002). To identify where the statistically significant differences exist in terms of mean population scores in the two tests after teaching, we will use the Tukey multiple comparisons method with a significance level of 5%.

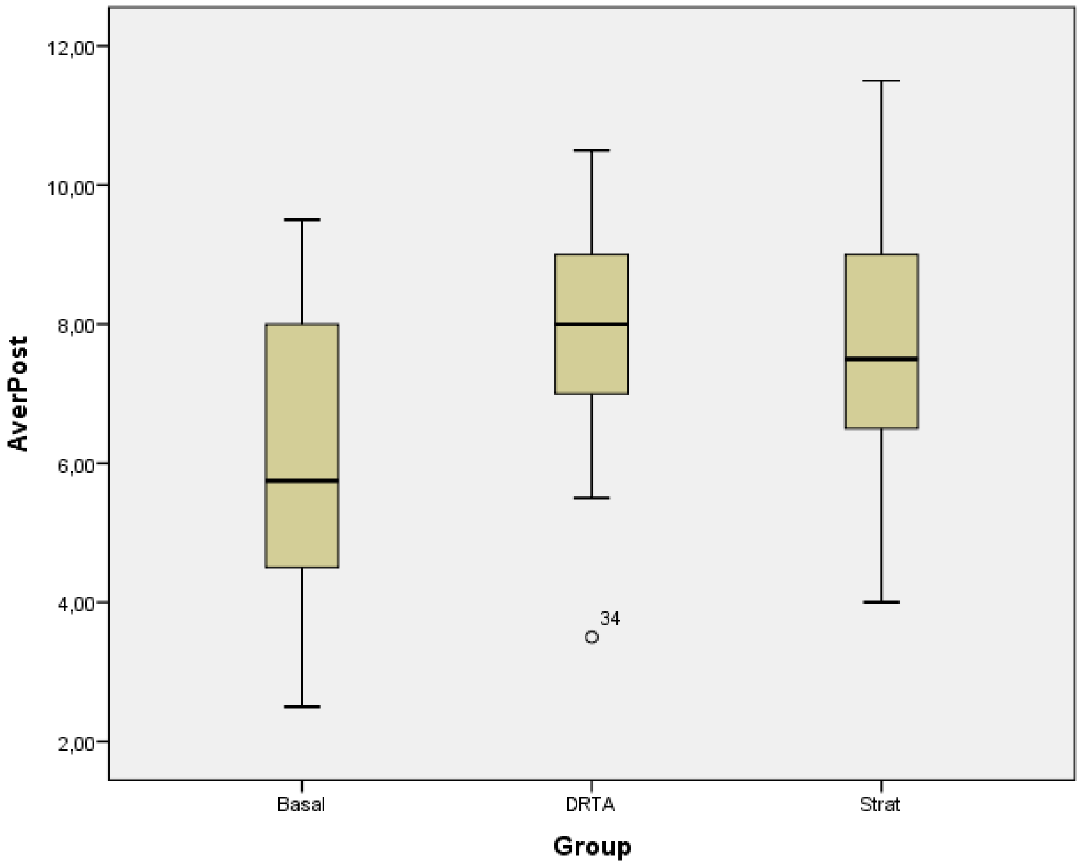
The mean score in the two tests with the 'Basal' teaching method differs statistically significantly from the other two, and it is worse, as indicated in the multiple comparisons table (Table 1).

Boxplot 1"



12,50
10,00
7,50
5,00
2,50
AverPost
52
34
Basal
DRTA
Strat
Group

Boxplot 2



12,00
10,00
8,00
6,00
4,00
2,00
AverPost
34
Basal
DRTA
Strat
Group

Boxplot 3

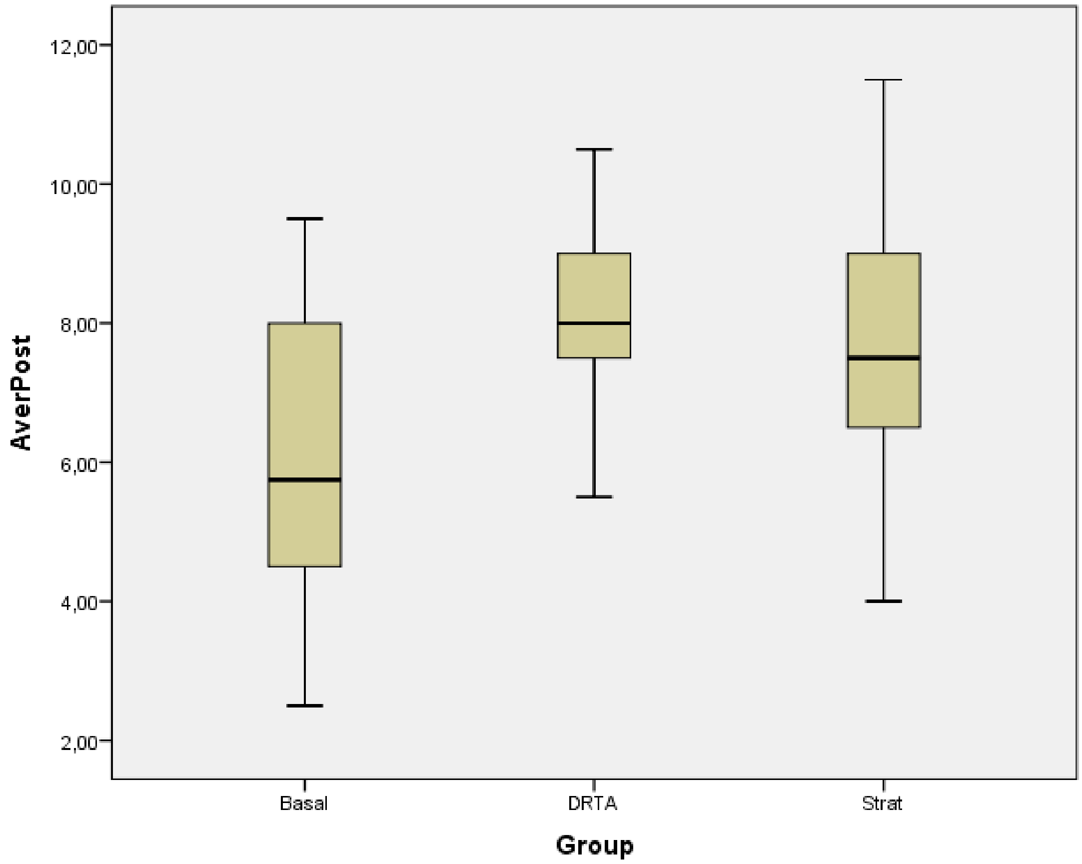


Table 1

	(I) Group	(J) Group	Mean Difference (I-J)	Sig.
Tamhane	Basal	DRTA	-2,10065*	,001
		Strat	-1,71970*	,029
	DRTA	Basal	2,10065*	,001
		Strat	,38095	,869
	Strat	Basal	1,71970*	,029
		DRTA	-,38095	,869