UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY,
BIOENGINEERING, ROBOTICS AND SYSTEM ENGINEERING

RESEARCH TRACK 2

Third Assignment

Statistic Report

Author:

Attolino Nicholas

Student ID:

s5656048

Professors:

Carmine Recchiuto

June 2, 2023

Contents

1 Introduction

In the third assignment of the Research Track 2 course, we were tasked with conducting a statistical analysis on the first assignment from the previous course, Research Track 1. The research entailed gathering data from two distinct implementations of the assignment and performing a statistical comparison between them. For this study, I examined both my personal implementation and my colleague Claudio Tomaiuolo's implementation. The task under analysis involves the collection of all silver tokens within the environment and positioning each silver token near a gold token but all the tokens are randomly distributed throughout the environment. My focus was on observing which of algorithms works best to be able to carry out statistical analysis on the speed of completion of the tasks.

2 Hypotheses made

In the realm of statistics, the "null hypothesis," denoted as H_0, represents the proposition that there is no noteworthy distinction between the two algorithms. Conversely, the "alternative hypothesis," referred to as H_a, asserts that a significant difference exists between the two implementations. Consequently, H_0 and H_a are mutually exclusive; accepting one necessitates the rejection of the opposing hypothesis. Below I mention null hypothesis and alternative hypothesis:

- Null Hypothesis: As a null hypothesis, both algorithms have performed effectively in Research Track 1 assignment 1. As a result, I assume that, on average, they have similar task completion speeds.
- Alternative Hypothesis: As an alternative hypothesis, I propose that due to the inherent differences between the two algorithms, they exhibit contrasting behaviors. Hence, one of the algorithms is expected to be faster than the other, not in every instance, but when observing their average performance.

3 Description and Motivation of the Experimental Setup

The primary objective of my experiment is to examine the performance of the two algorithms concerning task completion speed and compare these speeds to determine which algorithm exhibits better performance in average. To initiate the experiment, I modified the working environment of the robot. Initially, the environment consisted of 12 tokens, including 6 silver tokens and 6 gold tokens, with predefined positions. For the sake of efficiency, I accelerated the test by reducing the number of tokens by 6. In addition, to allow for the random location of each token in the environment, a part of the simulation's file *two-colours-assignment-arena.py* was modified with these line of code:

- (at the beginning of the code)

```
import random
```

- (into *def-init* function)

```
token.location = (random.uniform(-2, 2), random.uniform(-2, 2))
```

Whereas, a small piece of code was added to the 2 algorithms to calculate the execution time of performing the tasks:

- (before the *while*)

```
start-time = time.time()
```

- (into the *while*)

```
end-time = time.time()
elapsed-time = end-time - start-time
print("This is the elapsed-time: {}".format(elapsed-time))
```

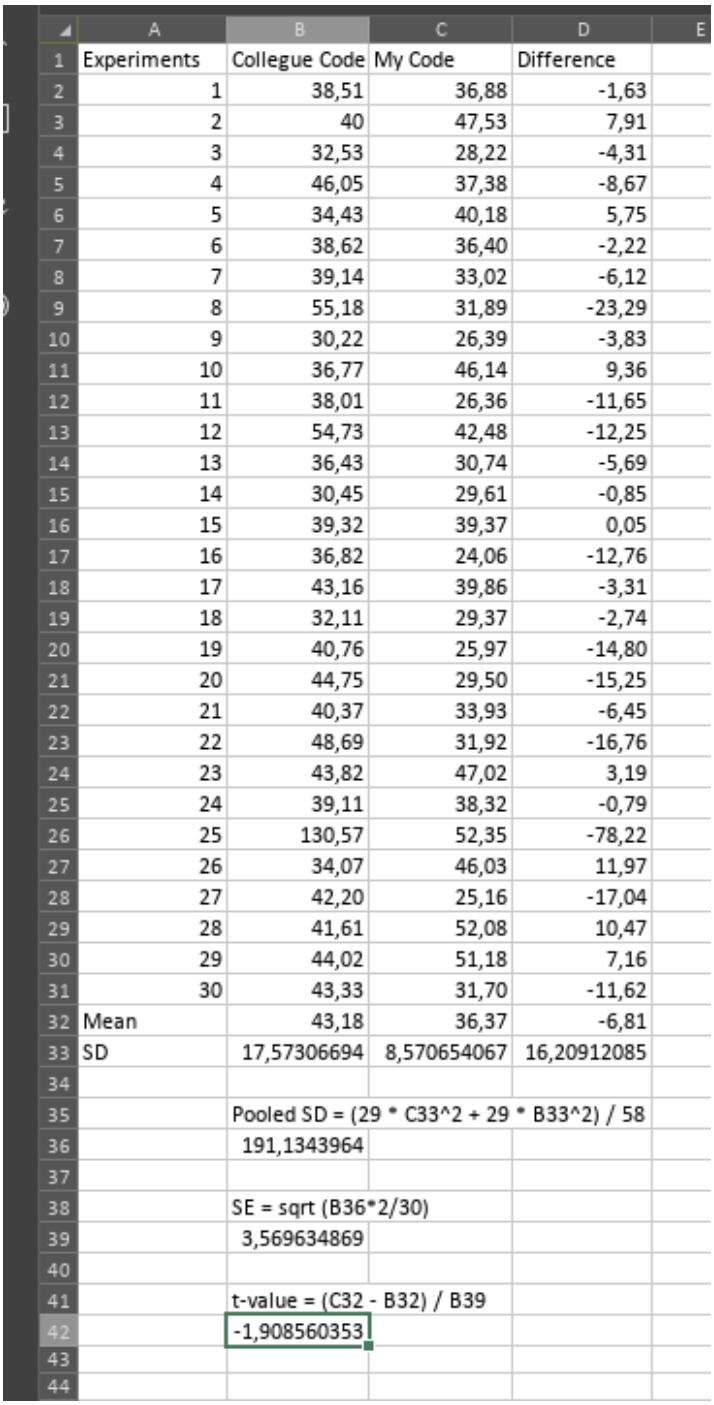
For conducting the statistical analysis, I employed the Paired T-Test, a statistical test utilized to compare the means of two populations when there are paired observations between the two samples. This test is highly valuable as it enables us to assess whether there exists a significant difference between the average values of the two measurement sets.

4 Results

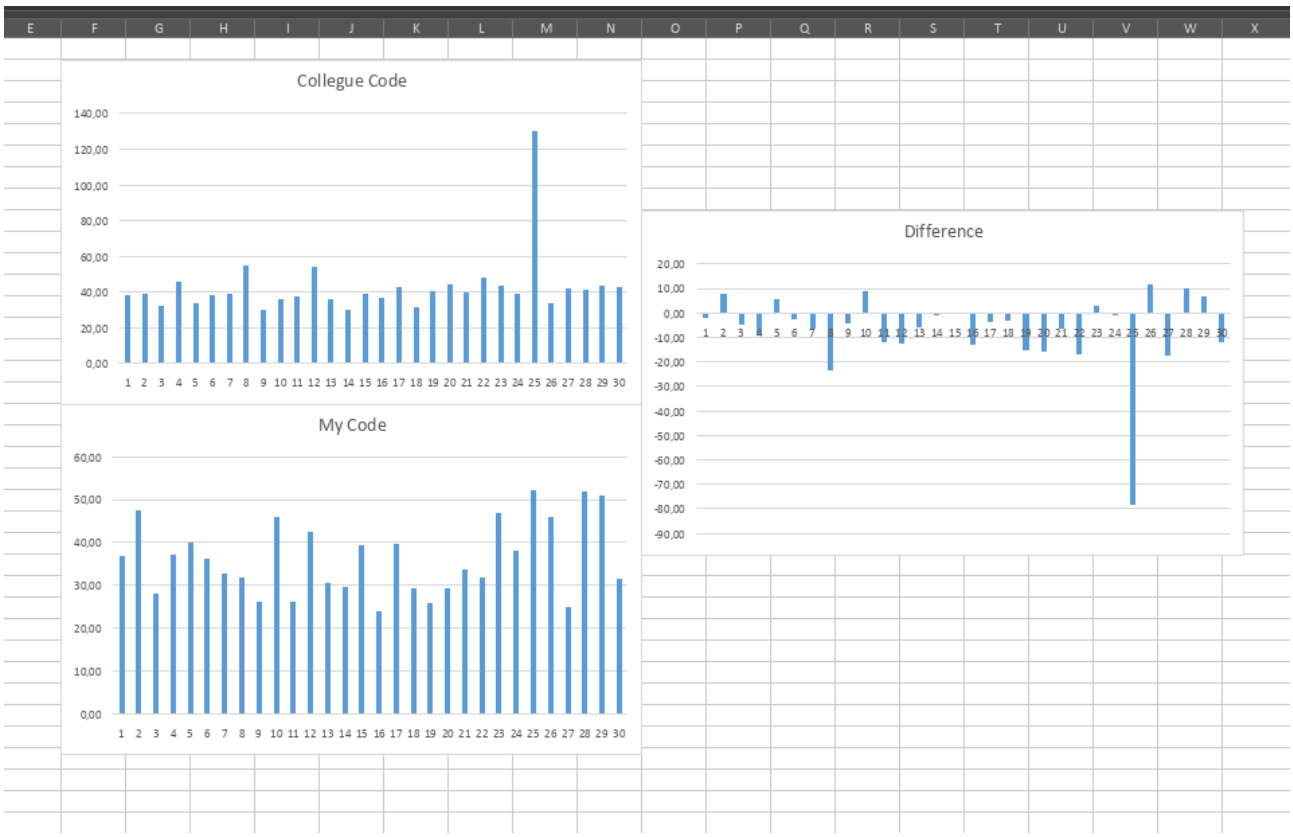
The algorithms were run 30 times, and below you can see the table with all the data obtained (and attached graphs). As it's possible to see in the table:

- the first column represents the number of experiments;
- the second and third columns represent the time to complete the tasks respectively of my colleague's algorithm and of mine;
- the fourth column contains all the computed differences between the second and the third columns.

In the image below is possible to see even the value of the mean of the differences, Standard Deviation, Standard Error and the t-value.



	A	B	C	D	E
1	Experiments	Collegue Code	My Code	Difference	
2	1	38,51	36,88	-1,63	
3	2	40	47,53	7,91	
4	3	32,53	28,22	-4,31	
5	4	46,05	37,38	-8,67	
6	5	34,43	40,18	5,75	
7	6	38,62	36,40	-2,22	
8	7	39,14	33,02	-6,12	
9	8	55,18	31,89	-23,29	
10	9	30,22	26,39	-3,83	
11	10	36,77	46,14	9,36	
12	11	38,01	26,36	-11,65	
13	12	54,73	42,48	-12,25	
14	13	36,43	30,74	-5,69	
15	14	30,45	29,61	-0,85	
16	15	39,32	39,37	0,05	
17	16	36,82	24,06	-12,76	
18	17	43,16	39,86	-3,31	
19	18	32,11	29,37	-2,74	
20	19	40,76	25,97	-14,80	
21	20	44,75	29,50	-15,25	
22	21	40,37	33,93	-6,45	
23	22	48,69	31,92	-16,76	
24	23	43,82	47,02	3,19	
25	24	39,11	38,32	-0,79	
26	25	130,57	52,35	-78,22	
27	26	34,07	46,03	11,97	
28	27	42,20	25,16	-17,04	
29	28	41,61	52,08	10,47	
30	29	44,02	51,18	7,16	
31	30	43,33	31,70	-11,62	
32	Mean	43,18	36,37	-6,81	
33	SD	17,57306694	8,570654067	16,20912085	
34					
35		Pooled SD = (29 * C33^2 + 29 * B33^2) / 58			
36		191,1343964			
37					
38		SE = sqrt (B36*2/30)			
39		3,569634869			
40					
41		t-value = (C32 - B32) / B39			
42		-1,908560353			
43					
44					



5 Discussion of the Results with Statistical Analysis

The procedure performed to achieve the goal is as follows:

1. The differences of each pair were calculated;
2. The mean difference was calculated;
3. The standard deviation of the differences was calculated;
4. The standard error of the mean difference was calculated;
5. The t-statistic was calculated;
6. The last step is using tables of the t-distribution to compare the value for T to the t(n-1) distribution; by doing that, we will give the p-value for the paired t-test.

	P						
one-tail	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	0.2	0.1	0.05	0.02	0.01	0.002	0.001
DF							
1	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	1.886	2.92	4.303	6.965	9.925	22.328	31.6
3	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965
18	1.33	1.734	2.101	2.552	2.878	3.61	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.85
21	1.323	1.721	2.08	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.5	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.06	2.485	2.787	3.45	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.689
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.66
30	1.31	1.697	2.042	2.457	2.75	3.385	3.646
60	1.296	1.671	2	2.39	2.66	3.232	3.46
120	1.289	1.658	1.98	2.358	2.617	3.16	3.373
1000	1.282	1.646	1.962	2.33	2.581	3.098	3.3
Inf	1.282	1.645	1.96	2.326	2.576	3.091	3.291

Figure 1: t-distribution table

6 Conclusion

Opting for an alternate hypothesis wherein one algorithm is quicker than the other on average, with no explicit indication of which algorithm is faster, I employed the one-tailed approach to scrutinize the data. For examination purposes, I examined the row associated with n-1, in this scenario, 29 (because there are 30 experiments - 1), and I selected a column as a point of reference, choosing the one containing the value that most closely approximates the previously computed t-statistic (1.908560353, the absolute value is considered). Subsequently, I investigated the corresponding value in the row representing the one-tailed analysis. As evident from the results, the corresponding value is: 0.025, signifying that I can embrace the alternative hypothesis with a 97.50% level of certainty. Ultimately, I can deduce that the null hypothesis is discarded, while the alternative hypothesis is accepted.