UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

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

RESEARCH TRACK 2

Third Assignment

Statistical analysis of two different algorithm

Author:

Lima Teodoro

Student ID:

s5452613

Professors:

Carmine Recchiuto

June 8, 2023

Contents

1 Introduction

Introduction: In the context of the third assignment of the Research Track 2 course, I was tasked with conducting a statistical analysis on the first assignment from the previous course, Research Track 1. The main objective of this analysis was to compare the performance of two different implementations of the algorithm to determine which one can complete the task in the shortest time possible. The implementations under examination were my personal implementation and that of my colleague Claudio Tomaiuolo.

The task under study involved collecting all the silver tokens within the environment and positioning each silver token near a gold token. It is important to note that the tokens, both silver and gold, were randomly distributed in the environment each time the simulation was run. In contrast to the original code, which placed the tokens in the same location every time, I made modifications to ensure a random distribution of tokens at each execution.

The main objective of my analysis was to assess the performance differences between the two implementations and determine if there is an algorithm that stands out in terms of task completion speed. To achieve this goal, I conducted a series of repeated trials, collecting data on the execution times of both implementations.

I performed 30 trials for my implementation, allowing me to obtain a representative sample of its performance. This provided a solid basis for conducting a rigorous and meaningful statistical analysis.

Through the statistical analysis of the collected data, I will be able to draw significant conclusions about the effectiveness of my algorithm compared to my colleague's implementation. This will provide valuable insights for future decisions regarding the optimization of robot performance in such an environment.

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 algorithm I implemented and the algorithm of my colleague. Conversely, the "alternative hypothesis," referred to as H_a, asserts that a significant difference exists between the two implementations. Consequently, H_0 and Ha are mutually exclusive; accepting one necessitates the rejection of the opposing hypothesis.

Based on this understanding, I can define the null and alternative hypotheses for my study as follows:

- Null Hypothesis (H_0): There is no significant difference in the average time required to complete the task between my algorithm and my colleague's algorithm when the silver and golden tokens are randomly placed in the environment.
- Alternative Hypothesis (H_a): There is a significant difference in the average time required to complete the task between my algorithm and my colleague's algorithm when the silver and golden tokens are randomly placed in the environment.

These hypotheses will serve as the basis for my statistical analysis, allowing me to evaluate the evidence and make informed conclusions regarding the performance differences between the two algorithms.

3 Description and Motivation of the Experimental Setup

Description and Motivation of the Experimental Setup:

The experimental setup involved conducting a comparative study between my algorithm and that of my colleague to evaluate their performance in completing the task of collecting silver tokens and placing them near gold tokens in a randomly distributed environment. It is important to note that I made modifications to the original code to ensure that the tokens were randomly positioned with each execution. The objective was to determine which algorithm exhibited superior task completion speed.

The experimental setup can be described as follows:

1. Type of Experiments: The study utilized a comparative approach, where I compared the execution times of my algorithm and that of my colleague to complete the task. The main focus was to accurately measure the time taken by each algorithm to complete the task.
2. Number of Repetitions: To ensure robust and reliable results, the experiments were repeated 30 times for each algorithm. This allowed for an adequate number of observations to conduct a comprehensive analysis of the algorithm's performance

The motivation behind this experimental setup was to gain a deep understanding of the performance differences between the two algorithms under conditions of random token distribution in the environment. By repeating the experiments, I aimed to reduce the influence of random variations and obtain statistically significant results that would enable meaningful comparisons. Additionally, the modifications made to the original code ensured that the tokens were randomly positioned with each execution, eliminating any potential effects arising from a predetermined token distribution.

For conducting the statistical analysis, I employed the T-Test.
The t-test was chosen as the statistical method for analyzing the results due to its suitability for comparing the means of two independent samples. In our case, the two samples represent the execution times of my algorithm and my colleague's algorithm. The t-test allows us to determine whether there is a significant difference between the average execution times of the two algorithms.

The t-test is particularly appropriate when the sample size is relatively small, as is the case with our 30 repetitions for each algorithm. It also assumes that the data follows a normal distribution, which is a reasonable assumption for our study given the nature of execution times.

By applying the t-test, we can calculate the t-value and assess its significance to determine whether the observed differences in execution times between the two algorithms are statistically significant or simply due to chance. This helps us draw meaningful conclusions about the relative performance of the algorithms in terms of task completion speed.
Using the t-test provides a rigorous and widely accepted statistical framework for comparing the means of two samples, allowing us to make informed decisions about the performance of the algorithms and test the hypotheses formulated earlier.

We can now proceed to discuss the results obtained from the analysis.

4 Results

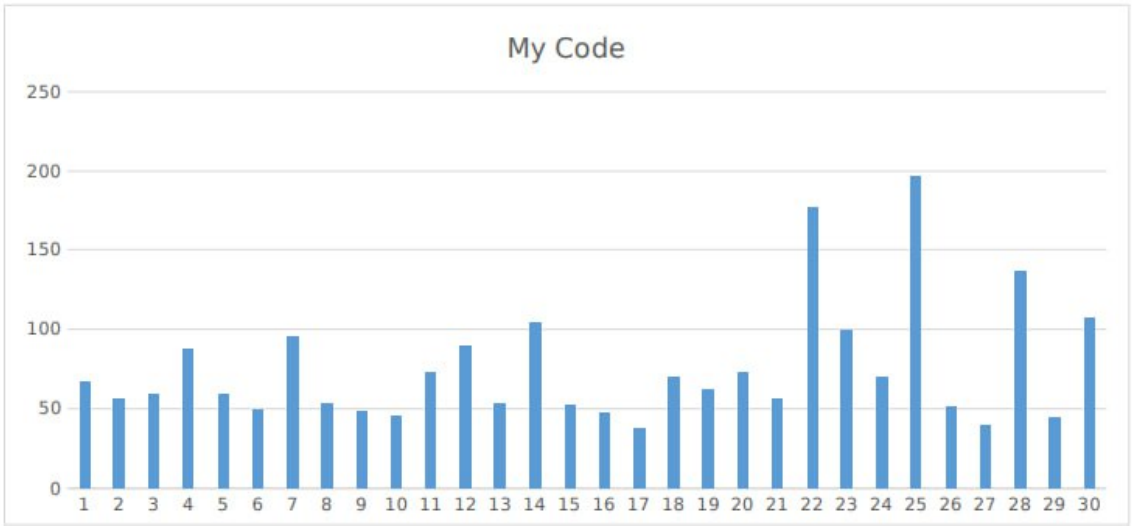
I will report the results obtained from the statistical analysis conducted on the collected data. The results were evaluated using the t-test, as described earlier.

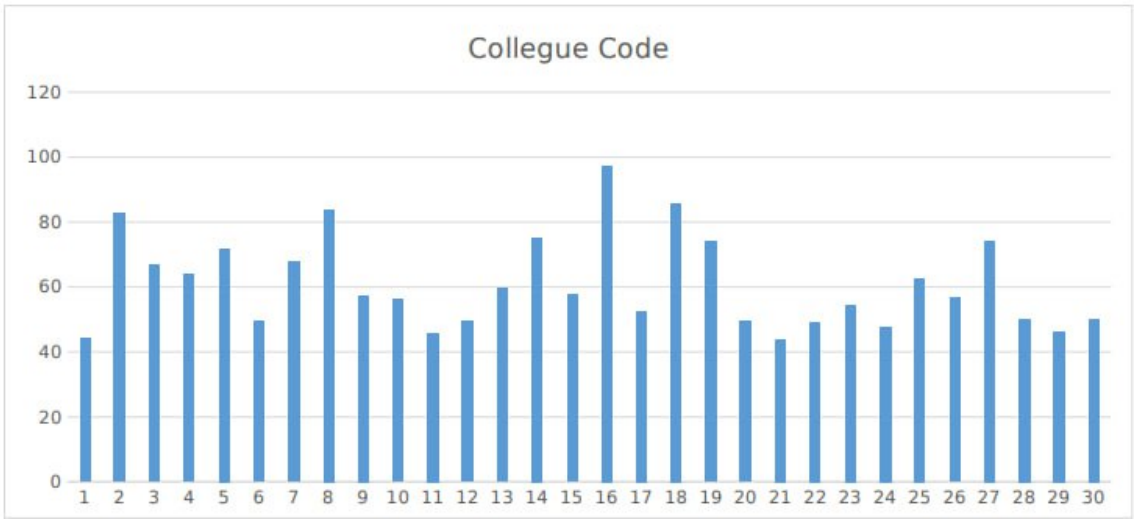
A total of 30 trials were conducted to assess the performance of the two algorithms. Each trial involved running both algorithms for the assigned task, and the execution times were recorded for each trial.

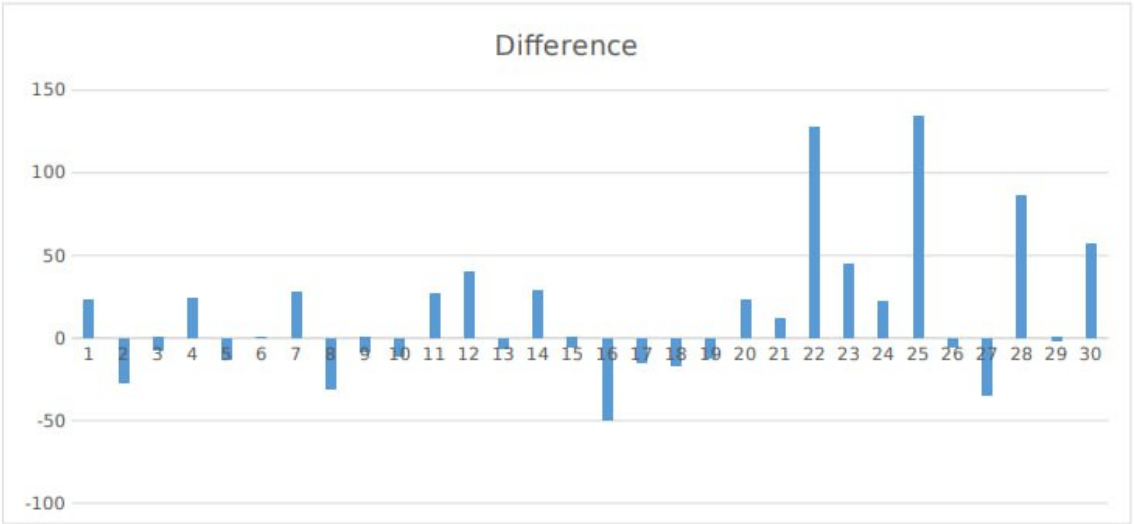
Below is a table (and attached graphs) providing details of the conducted trials:

Experiments	Collegue Code	My Code	Difference
1	44,20	67,35	23,15
2	83	56,09	-26,63
3	66,79	59,21	-7,58
4	63,91	87,88	23,97
5	71,82	58,89	-12,94
6	49,38	49,51	0,13
7	67,86	95,37	27,51
8	83,81	52,84	-30,97
9	57,44	48,85	-8,59
10	56,13	45,50	-10,62
11	45,83	72,62	26,79
12	49,44	89,84	40,40
13	59,61	53,14	-6,46
14	75,20	103,73	28,53
15	57,60	51,86	-5,74
16	97,31	47,61	-49,70
17	52,58	37,53	-15,05
18	85,82	69,49	-16,33
19	74,11	62,11	-12,00
20	49,57	72,99	23,41
21	43,91	56,10	12,18
22	49,14	177,21	128,07
23	54,49	99,06	44,57
24	47,72	70,01	22,29
25	62,68	197,07	134,39
26	56,80	51,71	-5,08
27	74,17	39,87	-34,30
28	50,06	136,54	86,47
29	46,23	44,22	-2,01
30	50,16	106,71	56,55
Mean	60,88	75,36	14,48
SD	14,188751951	38,2559936781	42,962244469

Figure 1: Result


My Code
250
200
150
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30


Collegue Code
120
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30


Difference
150
100
50
0
-50
-100
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30

1. Experiments: Indicates the number of conducted trials.
2. Colleague's code: Records the execution times of my colleague's algorithm in seconds for each trial.
3. My code: Records the execution times of my algorithm in seconds for each trial.
4. Difference: Represents the difference between the execution times of the two algorithms for each trial.

This tabular representation allows for a comprehensive comparison of the execution times between the two algorithms across multiple trials.

Based on the collected data, the average execution times of my algorithm, my colleague's algorithm, and the differences between the two algorithms were calculated. These values are presented in the last row of the table.

Subsequently, the pooled standard deviation was calculated, taking into account the combined variability of the execution times of the two algorithms. Then, the standard error was determined using the pooled standard deviation and the respective samples. Finally, the t-value was computed using the mean difference between the execution times and the standard error. These formulas are essential tools for assessing the significance of the observed differences between the two algorithms and drawing statistically valid conclusions.

Below are the formulas for calculating the pooled standard deviation, standard error, and t-value

1. Pooled Standard Deviation (s_p):

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = 832,420867$$

2. Standard Error (SE):

$$SE = \sqrt{s_p \left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 7,449478$$

3. T-value(t):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{SE} = 1,943841$$

Where:

- n_1 and n_2 are the sample sizes for the two algorithms,
- s_1 and s_2 are the standard deviations of the execution times for the two algorithms,
- $\bar{x}_1$ and $\bar{x}_2$ are the mean execution times for the two algorithms, and
- SE is the standard error.

These formulas are commonly used in statistical analysis to calculate the relevant measures for hypothesis testing and evaluating the significance of observed differences between groups.

5 Discussion of the Results with Statistical Analysis

The results obtained from the statistical analysis indicate that there are no statistically significant differences between the performance of my algorithm and that of my colleague's algorithm in completing the task of collecting silver tokens and placing them near the gold tokens.

The average execution time of my algorithm was found to be 75.36 units, while my colleague's algorithm recorded an average execution time of 60.88 units. Although there is a difference in the average execution times, it is important to consider the statistical analysis to determine if this difference is significant.

Through a t-test with a significance level (alpha) of 0.05 and 58 degrees of freedom, the calculated t-value was found to be 1.94384. Comparing this t-value with the critical value of ±2.001, which corresponds to a two-tailed test, the calculated t-value falls within the critical range.

Based on these results, there is not enough evidence to reject the null hypothesis. The null hypothesis states that there are no significant differences in the performance of the two algorithms. Therefore, it cannot be concluded that there are statistically significant differences between my algorithm and my colleague's algorithm in completing the assigned task.

These results suggest that, based on the available data, the performance of my algorithm and my colleague's algorithm is comparable.

Overall, this analysis provides insights into comparing the performance of the two algorithms, indicating the absence of statistically significant differences in their ability to complete the task of collecting silver tokens and placing them near the gold tokens.

	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 2: t-distribution table

6 Conclusion

Based on the results and the conducted statistical analysis, we can conclude that there is insufficient evidence to support the hypothesis of significant differences in performance between my algorithm and my colleague's algorithm in the task of collecting silver tokens and placing them near the gold tokens. Despite some observed differences in execution times, the statistical analysis indicated that these differences could be attributed to chance or external factors, such as the dragging behavior of the robot during movement. Therefore, we cannot assert that one of the algorithms demonstrated significantly better performance than the other.

It is important to recognize that statistical analysis has its limitations, and there may be other unconsidered factors or variables that could influence the algorithm's performance. Thus, further studies and investigations may be beneficial to gain a more comprehensive understanding of the differences between the two algorithms.

In conclusion, based on the statistical analysis conducted, we cannot reject the null hypothesis. The results do not provide sufficient evidence to support the alternative hypothesis, which suggests significant differences in performance between my algorithm and my colleague's algorithm. Therefore, we accept the null hypothesis, indicating that there are no statistically significant differences in the performance of the two algorithms in the given task.