UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

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

RESEARCH TRACK II

Third Assignment

Statistics

Author:

Mustafa Melih Toslak
Natnael Berhanu Takele

Student ID:

s5431021
s5446838

Professor:

Carmine Recchiuto

June 2, 2023

Contents

1 Introduction

In the context of the third assignment of Research Track II, our objective is to perform a comprehensive statistical analysis based on the initial assignment of Research Track I. This analysis specifically focused on comparing two distinct implementations: one developed by Natnael Berhanu Takele (S5446838) and the other one by Mustafa Melih Toslak (S5431021). For the purpose of clarity, Natnael's implementation will be referred to as ”N” while Melih's will be denoted as ”M” throughout this report.

The primary goal of the initial assignment in Research Track I was to develop a Python node that could effectively enable a robot to distinguish between silver and golden tokens, locate the nearest silver token, grasp it, identify the closest golden token, and release the silver token in proximity to the golden one. This process was to be repeated without reusing the same silver and golden tokens until no silver tokens remained.

The only difference from the first assignment in Research Track I is that the golden and silver tokens are distributed randomly in the same environment.

The subsequent sections of this report will present the experimental setup, the results obtained from the data analysis, and a discussion on the implications of the findings. Finally, we will draw conclusions based on the results and addressing the hypothesis.

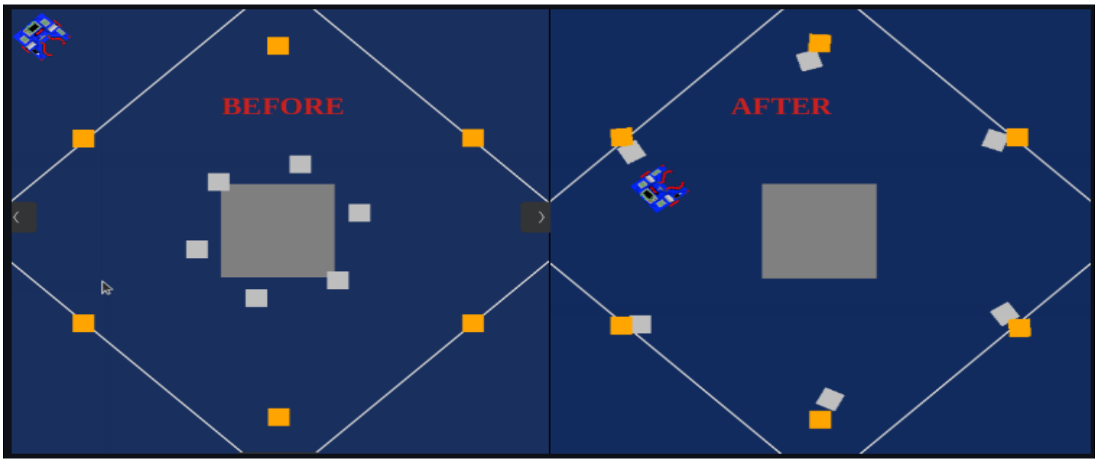

Figure 1: Research Track I Assignment 1

2 Hypotheses

The hypotheses for this study were formulated to investigate the performance difference between ”N” and ”M” algorithms in terms of completing the task. The hypotheses were tested using a T-test analysis on the completion time performance of each algorithm.

The null hypothesis (H0) posits that there is no significant difference between the performance of ”N” and ”M” algorithms in completing the task. In other words, any observed differences in completion time can be attributed to random chance or sampling variability.

On the other hand, the alternative hypothesis (HA) suggests that there is a specific difference between the performance of the two algorithms in completing the task. This hypothesis implies that the observed variations in completion time are not merely due to chance, but rather indicate a true disparity in performance between the algorithms.

To evaluate these hypotheses, a T-test analysis was conducted on the completion time data for each algorithm. The T-test provides a statistical measure to compare the means of two independent samples and assess whether the observed differences are statistically significant.

3 Experimental Setup

The experimental setup involved evaluating the performance of two algorithms, ”N” and ”M”, in completing a task. The task involved randomly placing silver and golden tokens in the environment, and the algorithms were tasked with finding and collecting these tokens. The experiments were conducted multiple times (40) to gather sufficient data for the analysis.

The experimental setup was designed as follows:

1. **Independent Testing**: Each algorithm was tested in isolation, without any interaction or interference between them. This ensured that the performance of each algorithm was evaluated independently.
2. **Number of Repetitions**: A total of 40 repetitions were conducted for each algorithm. This repetition allowed for a sufficient number of trials to gather reliable data for analysis.
3. **Recording of Failures**: In the event that an algorithm failed to complete the task in a specific repetition, it was recorded as a ”Failed” entry in the collected data.

By adhering to this experimental design, we aimed to gather comprehensive and meaningful data that would enable a thorough analysis and comparison of the performance between the two algorithms.

3.1 Random token distribution setup

In order to place the tokens randomly, ”robot-sim/sr/robot/arenas/two_colours_assignment_arena.py” file have been modified.

```
    import random

    INNER_CIRCLE_RADIUS = round(random.uniform(0.2,1), 1)   #It was 0.9
    OUTER_CIRCLE_RADIUS = round(random.uniform(1.5,2.4),1)  #It was 2.4
```

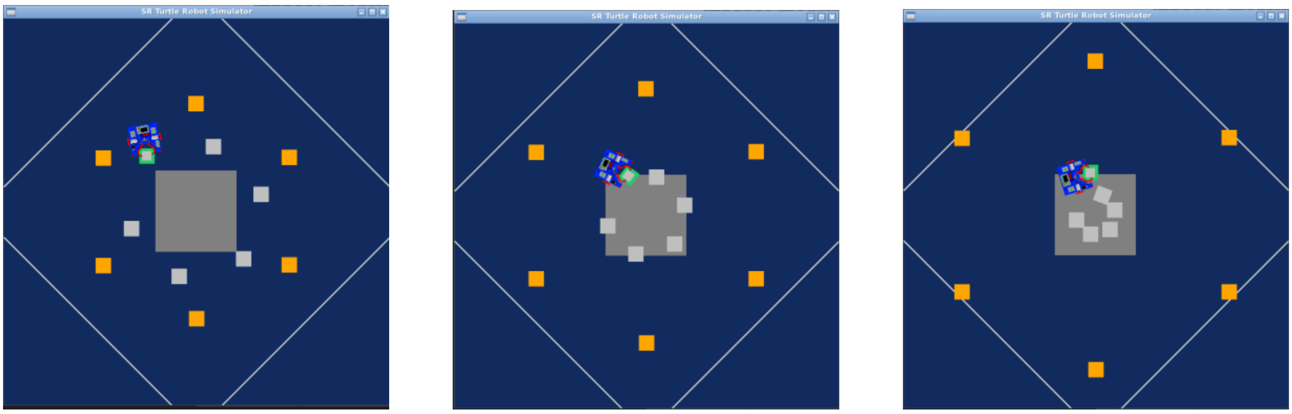

Figure 2: Randomly Distributed Tokens

3.2 Motor speed setup

In our experimental setup, it was ensured that the motor speeds used by both algorithms were identical.

```
    turn(+50, 0.1)
    turn(-50, 0.1)
    drive(80, 0.1)
elif rot_y < -a_th: turn(-5, 0.1)
elif rot_y > a_th:  turn(+5, 0.1)
```

3.3 Completion time setup

The main assignment code for the two algorithms have been modified to get the completion time of the task.

```
#Add this code before while loop
start_time = time.time()  # Start the timer

#Add these codes after the while loop
end_time = time.time()  # Stop the timer
elapsed_time = end_time - start_time
print("Elapsed Time:", elapsed_time, "seconds")
```

4 Experimental Results

In this section, the experimental results obtained from the performance analysis of "N" and "M" algorithms in terms of completing the task have been presented. The results are showed in the form of a graph, with the x-axis representing the repetitions and the y-axis denoting the completion time of each algorithm. The graph provides a clear visualization of the trends and patterns observed in the completion time data. Notably, the graph also indicates the occurrence of failed results, which are represented as blank areas on the graph.

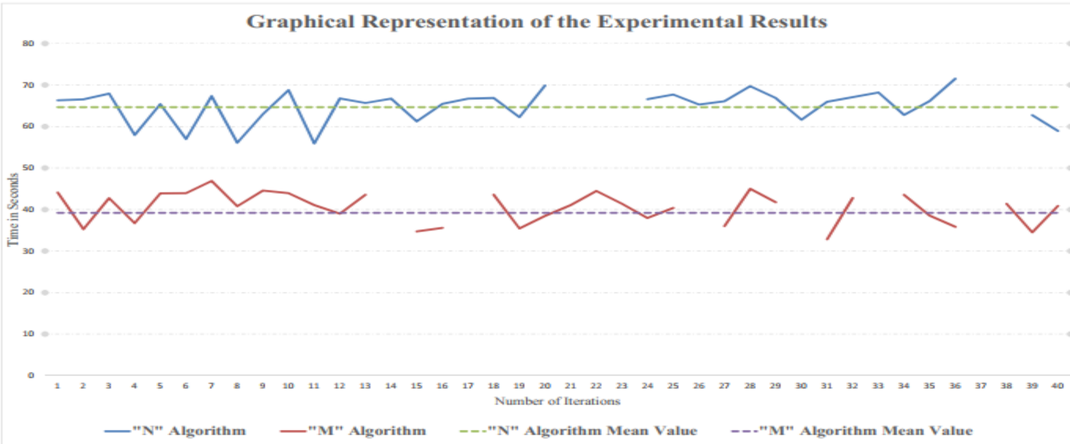

Figure 3: Graphical Comparison of the Algorithms

Upon examining the graph depicting the completion time data of "N" and "M" algorithms, several noteworthy observations can be made.

Firstly, it is evident that the completion time of the "N" algorithm tends to be higher compared to the "M" algorithm. This suggests that, on average, the first algorithm takes more time to complete the task compared to the second one.

Secondly, the presence of blank areas on the graph, representing failed results, provides insight into the algorithms' reliability. In this regard, it can be observed that the "N" algorithm exhibits a lower failure rate compared to the "M" algorithm.

To provide a comprehensive overview of the results, a detailed Excel table has been included in Appendix A. This table contains the recorded data for each algorithm, including the completion times and instances of failed results.

The experimental results serve as the basis for the subsequent statistical analysis, where a T-test will be employed to determine the significance of the observed differences in completion time. These results and the corresponding statistical analysis will be discussed in the following sections, enabling a deeper understanding of the relative performance of "N" and "M" algorithms in completing the assigned task.

5 Statistical Analysis ”T-Test”

In this study, a T-test analysis is conducted to assess the significance of the observed differences in completion time between ”N” and ”M” algorithms. The T-test is a statistical method that allows for the comparison of means between two independent samples. By applying this analysis to the completion time data of the two algorithms, we aimed to determine whether the differences in performance were statistically significant or merely due to chance.

The T-test assesses the null hypothesis, which assumes that there is no significant difference between the means of the two samples. The alternative hypothesis, on the other hand, posits that there is a specific difference between the means. By calculating the T-statistic and comparing it to the critical value from the T-distribution, we can determine whether to reject or fail to reject the null hypothesis.

The results of the T-test will provide valuable insights into the statistical significance of the observed differences in completion time between the ”N” and ”M” algorithms. These findings will contribute to a more robust understanding of the relative performance of the algorithms and help validate or refute the initial hypotheses.

5.1 Two-sample T-test

The two-sample t-test (also known as the independent samples t-test) is a method used to test whether the unknown population means of two groups are equal or not.

To perform the two-sample t-test, the average and the standard deviation values have been calculated.

Mean Value: $\mu_N = \frac{\sum_{n=1}^{N_N} X_i}{N_N} = 64.677$ and $\mu_M = \frac{\sum_{n=1}^{N_M} X_i}{N_M} = 39.220$

Standard Deviation: $\sigma_N = \sqrt{\frac{\sum_{n=1}^{N_N} (X_i - \mu_N)^2}{N_N}} = 4.017$ and $\sigma_M = \sqrt{\frac{\sum_{n=1}^{N_M} (X_i - \mu_M)^2}{N_M}} = 4.598$

Degree of Freedom: $DoF = N_N + N_M - 2 = 36 + 35 - 2 = 69$

Are the calculated mean values close enough for us to conclude that both algorithms perform in the same way? Or are the averages too different for us to make this conclusion?

Mean Value Difference : $\mu_N - \mu_M = 64.677 - 39.220 = 25.457$

This difference in our samples estimates the difference between the population means for the two algorithms.

Pooled standard deviation can be calculated. This builds a combined estimate of the overall standard deviation. The estimate adjusts for different group sizes.

During the experiment, a total of 40 iterations were conducted for each algorithm. However, it is important to note that there were 4 failed iterations for the ”N” algorithm and 5 failed iterations for the ”M” algorithm. These failures resulted in a decrease in the effective sample size for each algorithm.

For the statistical analysis, it is essential to consider the adjusted sample sizes to accurately assess the performance of the algorithms. The actual sample sizes will be taken into account when conducting further calculations and interpreting the results.

The pooled, estimated variance of the sampling distribution of the difference of means is then:

Pooled Variance: $\sigma_p^2 = \frac{(N_N - 1)\sigma_N^2 + (N_M - 1)\sigma_M^2}{N_N + N_M - 2} = 18.594$

Standard Error of Difference: $SED = \sqrt{\sigma_p^2 (\frac{1}{N_N} + \frac{1}{N_M}} = 1.011$

All the necessary components for calculating the test statistic have been determined. The calculation of the t-value can now be performed using the provided information and formulas.

T Value: $t = \frac{\mu_N - \mu_M}{SED} = 25.180$

With a significance level (α) of 0.05 and a degree of freedom of 69, we compared the calculated T-value of 25.180 to the critical T-value from the T-distribution table. The critical T-value for a two-tailed test at the given significance level is approximately ±1.994.

The absolute value of the calculated T-value (25.180) is significantly higher than the critical T-value (1.994), indicating a substantial difference between the performance of the ”N” and ”M” algorithms in terms of completing the task.

6 Conclusion

In conclusion, our statistical analysis using the t-test led us to reject the null hypothesis, providing strong evidence of a specific difference between the completion time performances of the ”N” and ”M” algorithms. The calculated T-value of 25.180 indicates a significant distinction, supporting the alternative hypothesis. Therefore, we can confidently state that the ”N” algorithm exhibits a significantly different completion time performance compared to the ”M” algorithm.

7 Appendix A

Experimental Results		
Number of Iterations	”N” Algorithm	”M” Algorithm
1	66.308	44.094
2	66.539	35.25
3	67.937	42.74
4	57.957	36.731
5	65.399	43.900
6	56.981	43.963
7	67.32	46.897
8	56.078	40.784
9	62.992	44.58
10	68.762	43.965
11	55.934	41.063
12	66.753	38.994
13	65.672	43.573
14	66.725	Failed
15	61.215	34.719
16	65.443	35.615
17	66.722	Failed
18	66.878	43.579
19	62.261	35.448
20	69.873	38.454
21	Failed	41.075
22	57.342	44.48
23	Failed	41.374
24	66.577	37.946
25	67.697	40.391
26	65.295	Failed
27	66.085	36.04
28	69.701	44.999
29	66.856	41.752
30	61.638	41.392
31	65.953	32.851
32	67.097	42.775
33	68.217	Failed
34	62.805	43.56
35	66.158	38.551
36	71.533	35.832
37	Failed	Failed
38	Failed	41.38
39	62.754	34.506
40	58.937	40.871

8 Appendix B

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

Figure 4: T-table

9 References

https://www.jmp.com/en_ch/statistics-knowledge-portal/t-test/two-sample-t-test.html
https://2022.aulaweb.unige.it/pluginfile.php/349272/mod_resource/content/0/Statistics.pdf
https://en.wikipedia.org/wiki/Student%27s_t-distribution