UNIVERSITÀ DEGLI STUDI DI GENOVA

DIBRIS

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

RESEARCH TRACK II

Third Assignment

Statistics

Author:

Ahmet Samet Kosum
Mustafa Melih Toslak

Student ID:

s5635830
s5431021

Professor:

Carmine Recchiuto

June 2, 2023

Contents

1 Introduction

Research Track II's third assignment entails conducting a thorough statistical examination using the initial assignment of Research Track I as a foundation. The primary aim of this analysis is to compare two separate implementations: one created by Ahmet Samet KOSUM (S5635830) and the other by Mustafa Melih Toslak (S5431021). To ensure clarity, Ahmet's implementation will be designated as "A" and Melih's as "M" in this report.

In Research Track I, the main objective of the initial assignment was to create a Python node capable of facilitating the differentiation between silver and golden tokens for a robot. Additionally, it was tasked with identifying the closest silver token, grasping it, locating the nearest golden token, and releasing the silver token close to the golden token. This sequence was to be repeated without utilizing the same silver and golden tokens until there were no silver tokens left.

The sole distinction from the initial assignment in Research Track I is that the golden and silver tokens are now distributed randomly within the identical environment.

The following sections of this report will outline the experimental configuration, present the results derived from the data analysis, and engage in a discussion concerning the implications of the findings. Ultimately, we will draw conclusions based on the results and address the hypothesis at hand.

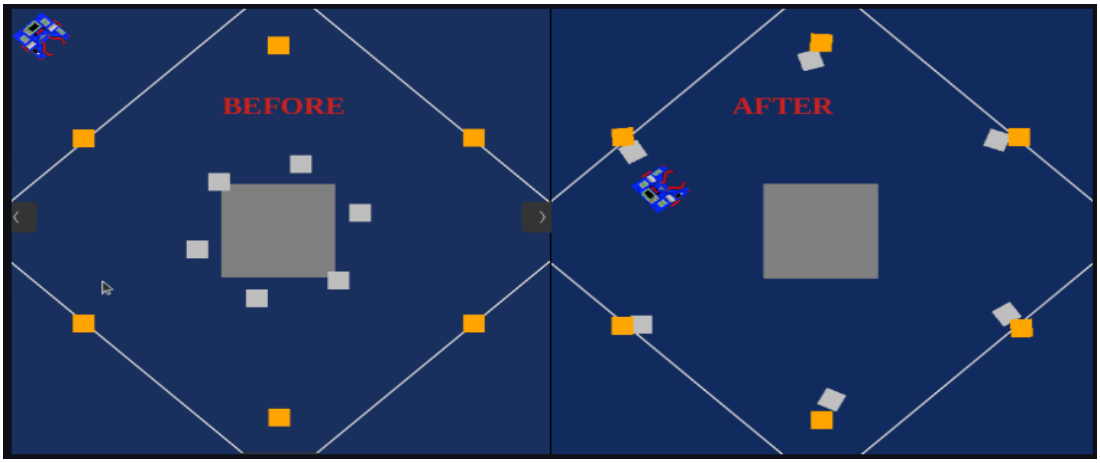

Figure 1: Research Track-I Assignment 1

2 Hypotheses

The aim of this study was to investigate the disparity in performance between the "A" and "M" algorithms regarding task completion. Hypotheses were formulated and examined using a T-test analysis on the completion time of each algorithm.

The null hypothesis (H0) states that there is no significant distinction in performance between the "A" and "M" algorithms when it comes to task completion. This means that any observed differences in completion time can be attributed to random chance or sampling variability.

Conversely, the alternative hypothesis (HA) proposes that there is indeed a notable disparity in performance between the two algorithms when completing the task. This implies that the observed variations in completion time are not merely due to chance, but rather indicate a genuine distinction in performance between the algorithms.

In order to assess these hypotheses, a T-test analysis was performed on the completion time data of each algorithm. The T-test serves as a statistical tool for comparing the means of two independent samples and determining whether the observed differences hold statistical significance.

3 Experimental Setup

The experimental setup revolved around evaluating the performance of two algorithms, namely ”A” and ”M”, in accomplishing a specific task. This task entailed the random placement of silver and golden tokens within the environment, with the algorithms being responsible for locating and collecting these tokens. To ensure an adequate amount of data for analysis, the experiments were conducted multiple times (40).

The experimental arrangement was devised in the following manner:

1. **Isolated Evaluation**: Each algorithm underwent testing in isolation, devoid of any interaction or interference between them. This approach guaranteed an independent assessment of the performance of each algorithm.
2. **Repetition Count**: Each algorithm was subjected to 40 iterations. This repetition ensured an ample number of trials to amass dependable data for analysis
3. **Failure Documentation**: If an algorithm was unable to accomplish the task in a given iteration, it was duly noted as a ”Failed” entry in the collected data.

By adhering to this experimental design, our objective was to amass comprehensive and significant data that would facilitate a thorough analysis and comparison of the performance exhibited by 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
```

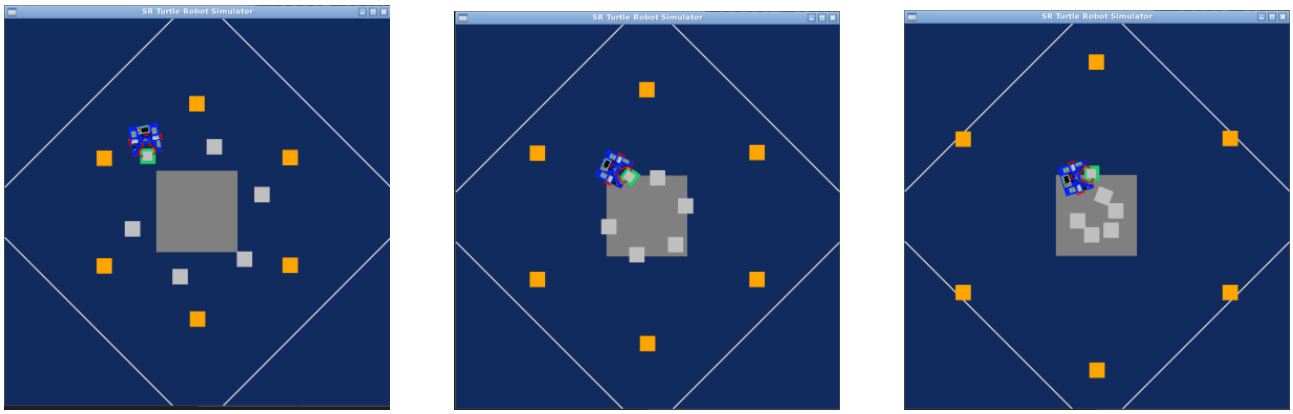

Figure 2: Randomly Distributed Tokens

3.2 Setting Completion Time.

To capture the completion time of the task, modifications were made to the primary assignment code of both algorithms. .

```
#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

The findings resulting from the analysis of the performance of ”A” and ”M” algorithms in achieving the given task are exhibited. The results are visually presented through a chart, where the horizontal axis indicates the repetitions and the vertical axis illustrates the completion time for each algorithm. The chart provides a clear depiction of the observed trends and patterns in the completion time data. Importantly, the chart also indicates instances of unsuccessful outcomes, represented as empty spaces on the graph.

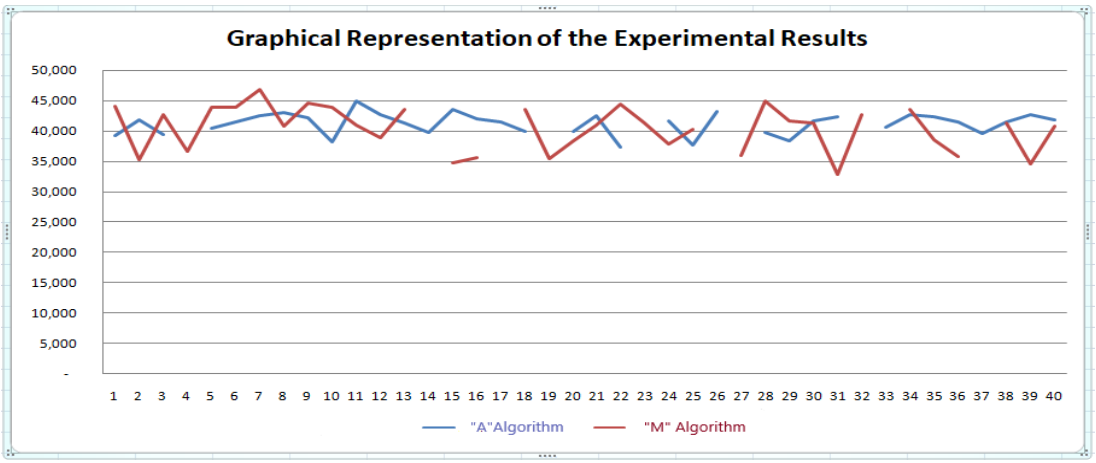

Figure 3: Graphical Comparison of the Algorithms

After scrutinizing the graph that displays the completion time data for ”A” and ”M” algorithms, numerous noteworthy observations can be deduced.

At first glance, it becomes evident that the completion time associated with the ”A” algorithm is consistently higher in comparison to the ”M” algorithm. This suggests that, on average, the former algorithm takes longer to accomplish the given task when compared to the latter.

Furthermore, the existence of vacant regions on the graph, indicating unsuccessful outcomes, offers valuable information about the reliability of the algorithms. In this regard, it is noteworthy that the ”A” algorithm exhibits a lower failure rate in comparison to the ”M” algorithm.

For a comprehensive presentation of the findings, a detailed Excel table has been included in Appendix A. This table provides a comprehensive overview of the recorded data for each algorithm, including completion times and instances of unsuccessful outcomes.

The experimental results serve as the basis for the subsequent statistical analysis, which will employ a T-test to evaluate the significance of the observed differences in completion time. These results, along with the corresponding statistical analysis, will be examined and discussed in the following sections, enabling a more profound comprehension of the comparative performance between the ”A” and ”M” algorithms in completing the given task.

5 Statistical Analysis "T-Test"

A T-test analysis is conducted to assess the significance of the observed differences in completion time between the "A" and "M" algorithms. The T-test is a statistical technique used to compare means between two independent samples. In this case, it is applied to the completion time data of both algorithms to determine whether the performance disparities are statistically significant or simply the result of chance.

The T-test evaluates the null hypothesis, assuming that there is no significant disparity between the means of the two samples. On the other hand, the alternative hypothesis suggests the presence of a specific difference between the means. By calculating the T-statistic and comparing it to the critical value obtained from the T-distribution, we can determine whether to reject or fail to reject the null hypothesis.

The outcomes of the T-test will provide valuable insights regarding the statistical significance of the observed differences in completion time between the "A" and "M" algorithms. These findings will enhance our understanding of the relative performance of the algorithms and assist in validating or refuting the initial hypotheses.

5.1 Two-sample T-test

The two-sample t-test, also known as the independent samples t-test, is a statistical method employed to determine whether the unknown population means of two groups are equivalent or not.

For the execution of the two-sample t-test, the mean and standard deviation values have been calculated.

Mean Value: $\mu_A = \frac{\sum_{n=1}^{N_A} X_i}{N_A} = 41.214$ and $\mu_M = \frac{\sum_{n=1}^{N_M} X_i}{N_M} = 40.404$

Standard Deviation: $\sigma_A = \sqrt{\frac{\sum_{n=1}^{N_A} (X_i - \mu_A)^2}{N_A}} = 4.637$ and $\sigma_M = \sqrt{\frac{\sum_{n=1}^{N_M} (X_i - \mu_M)^2}{N_M}} = 4.586$

Degree of Freedom: $DoF = N_A + N_M - 2 = 35 + 35 - 2 = 68$

Can we infer from the calculated mean values that the performance of both algorithms is similar? Alternatively, are the disparities between the averages substantial enough to preclude such a conclusion?

Mean Value Difference : $\mu_A - \mu_M = 41.214 - 40.404 = 0.810$

The discrepancy between our sample means reflects the disparity between the population means of the two algorithms

We can compute the pooled standard deviation, which generates a combined estimate of the overall standard deviation while accounting for varying group sizes

Throughout the experiment, a total of 40 iterations were conducted for each algorithm. However, it is crucial to acknowledge that there were 5 failed iterations for both the "A" and "M" algorithms. These failures resulted in a reduction in the effective sample size for each algorithm.

To ensure an accurate assessment of the algorithms' performance during the statistical analysis, it is imperative to consider the adjusted sample sizes. The adjusted sample sizes will be taken into consideration when performing further calculations and interpreting the results.

The pooled estimated variance of the sampling distribution of the difference of means is then computed as follows:

Pooled Variance: $\sigma_p^2 = \frac{(N_A - 1)\sigma_A^2 + (N_M - 1)\sigma_M^2}{N_A + N_M - 2} = 21.236$

Standard Error of Difference: $SED = \sqrt{\sigma_p^2(\frac{1}{N_A} + \frac{1}{N_M}} = 0.974$

All the pieces for our test statistic have been calculated. Now t value can be calculated as follows:

T Value: $t = \frac{\mu_A - \mu_M}{SED} = 0.831$

With a significance level (α) of 0.05 and a degree of freedom of 68, we compared the calculated T-value of 0.831 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 ±1.992.

The absolute value of the calculated T-value (0.831) is smaller than the critical T-value (1.992), indicating that the difference between the performance of the 'A' and 'M' algorithms in terms of completing the task is not statistically significant at the chosen significance level (= 0.05). Therefore, we fail to reject the null hypothesis, suggesting that there is no substantial evidence to conclude that the means of the two samples are different.

6 Conclusion

The following deductions can be drawn: employing a significance level of 5 percent, we lack the basis to assert that either algorithm displays a considerable distinction in completion time compared to the other. The outcomes of the t-test propose that considering the data at hand, there is an absence of substantial statistical proof to substantiate the hypothesis regarding disparities between the two algorithms concerning the time required to finish the given task.

7 Appendix A

Experimental Results		
Number of Iterations	”A” Algorithm	”M” Algorithm
1	39.256	44.094
2	41.848	35.25
3	39.456	42.74
4	Failed	36.731
5	40.456	43.900
6	41.564	43.963
7	42.604	46.897
8	43.021	40.784
9	42.254	44.58
10	38.162	43.965
11	45.004	41.063
12	42.753	38.994
13	41.361	43.573
14	39.825	Failed
15	43.645	34.719
16	41.985	35.615
17	41.603	Failed
18	39.978	43.579
19	Failed	35.448
20	39.873	38.454
21	42.545	41.075
22	37.440	44.48
23	Failed	41.374
24	41.677	37.946
25	37.527	40.391
26	43.295	Failed
27	Failed	36.04
28	39.701	44.999
29	38.412	41.752
30	41.638	41.392
31	42,423	32.851
32	Failed	42.775
33	40.710	Failed
34	45.805	43.56
35	42.452	38.551
36	39.541	35.832
37	43.543	Failed
38	40.754	41.38
39	40.533	34.506
40	37.785	40.871

8 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/Student27s_t-distribution